# Best Available Copy

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

## CORRECTED VERSION

(19) World Intellectual Property Organization International Bureau



## 10 DE 1800 E 10 DE 1800 E 1800 E

(43) International Publication Date 6 September 2002 (06.09.2002)

PCT

(10) International Publication Number WO 02/068666 A1

- (51) International Patent Classification7:
- \_\_\_\_
- •
- (21) International Application Number: PCI/US01/43588

C12N 15/82

- (22) International Filing Date:
  - 16 November 2001 (16.11.2001)
- (25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/249,608

17 November 2000 (17.11.2000) US

- (71) Applicant (for all designated States except US): GENEN-COR INTERNATIONAL, INC. [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): DUNN-COLE-MAN, Nigel [US/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). LANGDON, Timothy [GB/GB]; Troedyrhiw, Lon Tyllwyd, Llanfarian, Aberystwyth SY23 4UJ (GB). MORRIS, Phillip [GB/GB]; 6 Trefaenor, Comins-coch, Aberystwyth SY23 3UB, Cerdigion, Wales (GB).
- (74) Agent: BOYD, Victoria, L.; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).

- (81) Designated States (national): AE, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, EB (utility model), EB, ES, FI (utility model), FI, GB, GL, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BI, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

#### Published:

- with international search report
- (48) Date of publication of this corrected version: 6 November 2003
- (15) Information about Correction: see PCT Gazette No. 45/2003 of 6 November 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

(57) Abstract: Described herein are methods to enhance the production of more highly fermentable carbohydrates in plannts, especially forage grasses. The invention provides for transgenic plants transformed with expression vectors containing a DNA sequence encoding ferulic acid esterase I from Aspergillus, preferably A. Niger. The expression vectors may optionally comprise a DNA sequence encoding xylanase from Trichoderma, preferably T. reesei. Expression of the enzyme(s) is targeted to specific cellular compartments, in specific tissues and under specific environmental conditions. Uses of this invention include, but are not limited to, forage with improved digestibility for livestock, and enhanced biomass conversion.



O 02/068666 A

5

# MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

Inventors: Nigel Dunn-Coleman Timothy Langdon Phillip Morse

15

25

30

35

10

## CROSS-REFERENCE TO RELATED APPLICATIONS

Pursuant to 35 U.S.C. §119(e), the present application claims benefit of and priority to USSN 60/249,608, entitled "MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF FORAGE GRASS CELL WALLS BY TARGETED EXPRESSION OF A FERULIC ACID ESTERASE GENE", filed November 17, 2000, by Morris et al.

## FIELD OF THE INVENTION

This invention relates to methods to enhance to availability of fermentable carbohydrates.

## **BACKGROUND OF THE INVENTION**

The present crisis in livestock agriculture has prompted a resurgence of interest in grass-fed animals. However, while a high-forage diet may be desirable, it does not currently satisfy the demands of modern animal production. For the animal to make efficient use of the forage it consumes, the energy demands of the microorganisms in the rumen must be met and synchronized with the availability of plant proteins. Otherwise this lack of synchrony will lead to (a) proteins and other nutrients being poorly utilized in the rumen, (b) loss of nitrogen, in urine and feces and therefore, the environment and (c) the need to feed excessive amounts of protein concentrates as supplements to the ruminant diet.

Cellulose and hemicellulose in grass and maize tissues could meet the

energy requirements of the ruminant or provide new feed-stocks for industrial fermentation to ethanol. This potential is not currently realized because the cell walls are lignified and the cell wall polysaccharides highly cross-linked with phenolic residues and lignin, resulting in low rates of plant cell wall digestion in comparison to rates of protein breakdown in ruminants. This is a particular problem for the most important forages in Europe, the ryegrasses *Lolium* perenne and *L. mutiflorum* as well as one of the major impediments to the wider use of better adapted species, such as *Festuca arundinacea*, as a forage crop. Increasing the digestibility index of grasses has therefore been a major breeding objective for several decades but progress has been slow due to difficulties in fixing natural variation in the synthetic varieties derived from these outbreeding species (Hayward, et al., TAG 70:48 (1985)).

10

15

20

25

Removing labile phenolics by chemical treatment with alkali is known to increase the biodegradability and nutritional value of low-quality feed such as cereal straw, and is employed commercially for feed upgrading. Reducing phenolic cross-linking of cell wall carbohydrates is therefore a predictable way of improving the rate of digestion and digestibility of ryegrass. However chemical modification may have other disadvantages. Therefore, genetic modification would be a preferable method of changing the cell wall chemistry of highly digestible varieties. Many in the field are pursuing this approach. An alternative, however, is to use genetic modification to reduce the levels of phenolic acids in the cell walls available for crosslinking either by directly disrupting ester bonds linking phenolics and lignins to cell wall polysaccharides or by preventing excessive ferulation of cell wall carbohydrates prior to their incorporation into the cell wall.

This invention meets this and other needs by using targeted or inducible expression of cell-wall-degrading enzymes in plants.

10

15

20

25

## SUMMARY OF THE INVENTION

Provided herein are methods for enhancing the availability of fermentable carbohydrates. In one aspect, there is provided an expression cassette comprising a DNA sequence encoding at least one cell wall degrading enzyme. The DNA sequence encoding at least one cell wall degrading enzyme may be operatively linked to a promoter sequence. The promoter may be constitutive or inducible. The expression cassette may further comprise a targeting sequence.

In one embodiment, the cell wall degrading enzyme is selected from the group consisting of ferulic acid esterase, xylanase, xylosidase, cellulase, endoglucanase, and cellbiohydrolase. In a preferred embodiment cell wall degrading enzyme is derived from a fungal source. In a more preferred embodiment, the fungal ferulic acid esterase is an Aspergillus ferulic acid esterase, preferably A. niger. In another embodiment the xylanase is derived from Trichoderma, preferably T. reesei.

In another aspect of the invention, there is provided a plant transformed with the expression cassette comprising a DNA sequence encoding at least one cell degrading enzyme. The plant may be selected from the group consisting of Festuca, Lolium, Avena and Zea. In a preferred embodiment the plant is a forage grass. In another embodiment, the plant is maize.

Further provided herein is a method of controlling the level of phenolic acids in plant cell walls of a transgenic plant. The method, in one embodiment, comprises introducing to a plant cell an expression cassette comprising a DNA sequence encoding at least one cell wall degrading enzyme, preferably a ferulic acid esterase.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.

## BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates a restriction map of a DNA fragment containing the gene encoding the 38kd ferulic acid esterase.

Figures 2 A-E illustrate the complete DNA (SEQ. ID NO:\_\_), with highlighting to point out the signal sequence, intron and various restriction endonuclease sites, and amino acid sequence (SEQ. ID. NO:\_\_) corresponding to the gene encoding the 38 kD ferulic acid esterase isolated from *Aspergillus niger*.

Figure 3 illustrates the DNA sequence of the gene encoding the 38 kD esterase (SEQ. ID. NO:\_\_).

Figure 4 Illustrates the construction of the Intronless ferulic acid esterase isolated from Aspergillus niger.

Figure 5 illustrates that the overlapping of PCR products made with primers FAE-I5 FAE-I3 creates two possible uninterrupted reading frames — the top in the figure below is functional (highlighted serine is at active site), the bottom is inactivated.

Figure 6 illustrates the possible vector constructions useful in the present invention. Various combinations are possible. Although and FAE gene is depicted another cell wall degrading enzyme may be used alone (i.e., instead of) or in conjunction with the FAE gene. Amp = ampicillin resistance gene.

Figure 7 illustrates pCOR105.

20

25

Figure 8 illustrates a generic ALE-TER vector.

Figure 9 illustrates the KDEL-COOH ER retention sequences.

Figure 10 illustrates the FAE-LINKER-FRAMESHIFT structure and sequence.

Figure 11 illustrates plant transformation cassettes.

Figure 12 is a table of the vectors used herein.

Figure 13 depicts the barley aleurain vacuolar and apoplast signal sequence.

Figure 14 illustrates the rat sialyl transferase structure and sequence.

Figure 15 illustrates the potato protease inhibitor II (PPI) motif structure and sequence.

Figure 16 illustrates the targeted expression of gfp to different cell compartment. Also shown are schematics of the vectors used.

5

10

15

20

25

30

Figure 17 illustrates the FAE activity in transgenic Festuca arundinacea leaves of different ages under ER and APO targeting sequences.

Figure 18 illustrates the FAE activity in transgenic Festuca arundinacea leaves of different ages under Vac targeting sequence.

Figure 19 illustrates the FAE activity in transgenic Lolium mutflorum leaves of different ages.

Figure 20 illustrates the FAE activity in transgenic Lolium mutilorum leaves under Vac, ER and APO targeting sequences.

Figure 21 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under Vac targeting sequence.

Figure 22 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under APO and ER targeting sequence.

Figure 23 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature Festuca arundinacea plants expressing FAE under an actin promoter.

Figure 24 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Lolium mutflorum* plants expressing FAE under an actin promoter.

Figure 25 illustrates the rate of fermentation and cumulative gas production in Festuca arundinacea cells.

Figure 26 illustrates the in vitro fermentation of Festuca arundinacea cell walls from cell cultures expressing recombinant FAE1.

Figure 27 illustrates the Time to maximum rate digestion for Festuca arundinacea cells.

Figure 28 illustrates the total gas production in Festuca arundinacea cells.

Figure 29 illustrates the kinetics of FAE activity by ferulic acid release from cell wall under self digestion in Festuca arundinacea and stimulation by xylanase.

Figure 30 illustrates the beta-glucoronidase activity under the Lolium See1 senescence promoter in leaves of transgenic plants of *Lolium mutflorum*.

Figure 31 illustrates the release of monomeric and dimeric HCAs on self digestion of leaves of vacuolar targeted FAE expressing plants.

Figure 32 is a schematic of the pTP10-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

5

10

15

20

Figure 33 is a schematic of the pUA4-4 vector. Also shown is the 5345 bp nucleotide sequence of the vector.

Figure 34 is a schematic of the pTU4 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 35 is a schematic of the pTT5.14 vector. Also shown is the 5395 bp nucleotide sequence of the vector.

Figure 36 is a schematic of the pTP8-5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 37 is a schematic of the pTP5-1 vector. Also shown is the 5277 bp nucleotide sequence of the vector.

Figure 38 is a schematic of the pTP4a2 vector. Also shown is the 5327 bp nucleotide sequence of the vector.

Figure 39 is a schematic of the pTP3-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

Figure 40 is a schematic of the pTU5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 41 is a schematic of the pGT6 vector. Also shown is the 4773 bp nucleotide sequence of the vector.

Figure 42 is a schematic of the pJQ5 vector. Also shown is the 5034 bp nucleotide sequence of the vector.

25 Figure 43 is a schematic of the pJO6.1 vector. Also shown is the 4950 bp nucleotide sequence of the vector.

Figure 44 is a schematic of the pJQ4 vector. Also shown is the 4974 bp nucleotide sequence of the vector.

Figure 45 is a schematic of the pPQ10.1 vector. Also shown is the 5164 bp nucleotide sequence of the vector.

Figure 46 is a schematic of the pJQ3 vector. Also shown is the 4965 bp nucleotide sequence of the vector.

Figure 47 is a schematic of the pUG4 vector. Also shown is the 5295 bp nucleotide sequence of the vector.

Figure 48 is a schematic of the pUB8.11 vector. Also shown is the 5001 bp nucleotide sequence of the vector.

Figure 49 is a schematic of the pTP11-1 vector. Also shown is the 5387 bp nucleotide sequence of the vector.

Figure 50 illustrates the actin promoter and its corresponding nucleotide sequence.

Figure 51 illustrates the Aleurain-NPIR delete structure. The corresponding nucleotide sequences are also shown.

Figure 52 Illustrates the SEE1 (senescence enhanced) promoter sequence.

Figure 53 illustrates the SEE1 (senescence enhanced) promoter sequence plus the vacuolar aleurain signal/NPIR sequence.

15

20

10

5

## **DETAILED DESCRIPTION OF THE INVENTION**

The invention will now be described in detail by way of reference only using the following definitions and examples. All patents and publications, including all sequences disclosed within such patents and publications, referred to herein are expressly incorporated by reference.

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, et al., Dictionary of Microbiology and Molecular Biology, 2D ED., John Wiley and Sons, New York (1994), and Hale & Marham, The Harper Collins Dictionary of Biology, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy

orientation, respectively. Practitioners are particularly directed to Sambrook et al., 1989, and Ausubel FM et al., 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

#### Definitions

5

10

15

25

30

It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an, "and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise.

"Conservatively modified variants" applies to both amino acid sequences and polynucleotides. With respect to particular polynucleotides, conservatively modified variants refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical polynucleotides encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every polynucleotide herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a polynucleotide (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a polynucleotide which encodes a polypeptide is implicit in each described

sequence. For purposes of protein expression, there are "sub-optimal codons." These are codons that are not preferred by a particular genus or species. Altering these "sub-optimal codons" to "preferred codons" is a silent mutation in that the amino acid encoded by the codons is the same but one codon is preferentially expressed by the particular genus, e.g., Triticum spp.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a polynucleotide, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);

10

15

20

30

of the plant.

- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

(see, e.g., Creighton, Proteins (1984)).

"Pyroglutamic acid" is the cyclized internal amide of L-glutamic acid
The phrase "controlling the level of phenolic acids" refers to the
manipulation of phenolic acid expression in plants, particularly plant cell walls.
The manipulation can be either positive; e.g., increasing the levels of phenolic
acids; negative, e.g., decreasing the level of phenolic acids; or neutral, e.g.,
changing the relative amounts of specific phenolic acids in the cell walls but
keeping the total amount relatively the same. The timing of manipulation can be
during plant growth or after plant growth, e.g., after a plant has been cut or pulled
from the ground or ingested. "Plant cell walls" refers to the cell walls of any cell

The term "derived" means that a polynucleotide or protein is related to

another polynucleotide or protein. The relations can be one of homology, e.g., nucleotides and proteins from certain species are homologous to similar polynucleotides and proteins of other species; analogy, e.g., proteins perform the same function and therefore are related to each other regardless of organism of origin. The relationship can be a man-made one, e.g., a protein (and a polynucleotide) can be derived from another protein by mutation; or chemical manipulation (peptidomimetics). Furthermore, a protein or a polynucleotide can be derived from an organism if, in the natural state, the protein or polynucleotide is found in one organism but recombinantly produced in another.

The term "exogenous polynucleotide" refers to a polynucleotide which is introduced into the plant by any means other than a sexual cross or sexual reproduction. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous polynucleotide is referred to here as an R<sub>1</sub> generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are progeny of such a plant.

10

15

20

30

The term "isolated polynucleotide molecule" or "isolated protein" refers to a polynucleotide or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *FAE1* gene is separated from open reading frames which flank the gene and encode a protein other than FAE1.

A "FAE1 encoding polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of an FAE1 gene or which encodes a FAE1 polypeptide. FAE1 polynucleotides can also be identified by their ability to hybridize under low stringency conditions (see below) to nucleic acid probes having a sequence of 8 to 300 bases, preferably a sequence of 80 to 100 bases

in the sequence disclosed in WO 98/14594.

10

15

20

25

30

The term "nucleic acid encoding," "nucleic acid sequence encoding" or "polynucleotide encoding" refers to a polynucleotide which directs the expression of a specific protein or peptide. The polynucleotides include both the DNA strand sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The polynucleotides include both full length polynucleotides as well as shorter sequences derived from the full length sequences. It is understood that a particular polynucleotide includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in a specific host cell. The polynucleotide includes both the sense and antisense strands as either individual single strands or in the duplex form.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

The term "plasmid" refers to a circular double stranded DNA molecule which comprises the coding sequence of interest, regulatory elements, a selection marker and optionally an amplification marker. A plasmid can transform prokaryotic cells or transfect eukaryotic cells. An "expression cassette" means a portion of a plasmid (or the entire plasmid) containing the regulatory elements desired for transcription, translation and/or expression and the coding region of a polynucleotide. A plasmid may contain one or more expression cassettes. If multiple expression cassettes are introduced into a plant, they can be introduced simultaneously or at different times. If simultaneous introduction is desired, the expression cassettes can be on one plasmid or more. Typically, an expression cassette comprises a promoter, poly A+ tail, and signal sequences that target the expressed polypeptide to a specific region of a cell or to be secreted, if desired. Examples of signal sequences that "target expression" of ferulic acid esterase include sequences located upstream of the FAE coding sequence. The polynucleotide that encodes the signal sequence is found preferably within the 100 nucleotides "upstream" (in the 5' direction) from the initiation codon (AUG). More preferably, the polynucleotide that encodes the signal sequence is found within the 50 nucleotides upstream from the initiation

codon. Many different cellular organelles are targeted by the signal sequences used in this invention. The organelles include, but are not limited to, vacuoles, Golgi apparati, endoplasmic reticula, and apoplasts. In addition to upstream signal sequences, the expression cassette of this invention may include a polynucleotide that encodes a signal sequence at the 3' end. These signal sequences include, but are not limited to stop codons and the KDEL sequence. In addition to KDEL, other similar sequences are contemplated by this invention, including but not limited to RDEL. In addition to a KDEL sequence, a signal sequence can include a linker to a KDEL sequence. A linker is an extension of the reading frame of the encoding polynucleotide to the signal sequence. Preferably, the polynucleotide encoding the signal sequence is directly downstream from the coding sequence, more preferably less than 100 base pairs from the stop codon, more preferably less than 20 base pairs from the stop codon.

10

15

20

25

30

The term "polynucleotide," "polynucleotide" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses polynucleotides containing known analogues of natural nucleotides which have similar binding properties as the reference polynucleotide and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular FAE1 polynucleotide of this invention also implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); and Cassol et al., 1992; Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)). The term polynucleotide is used interchangeably with gene, cDNA, and mRNA encoded by a gene.

The term "polypeptide," "peptide," and "protein" are used interchangeably and refer to amino acids connected by peptide bonds. Polypeptides can be

entire proteins or portions thereof. For Example. a FAE1 polypeptide may refer to the entire FAE1 protein or fragments of the FAE1 protein. A "ferulic acid esterase with an altered glycosylation site" refers to a FAE protein wherein a mutation has changed the glycosylation pattern of the protein. Mutations that effect such changes are well known in the art and include, but are not limited to, amino acid substitutions, and mutations in the proteins of the Golgi apparatus and endoplasmic reticulum that effect glycosylation of proteins.

The term "promoter" refers to a polynucleotide that directs expression of a coding sequence. A promoter can be constitutive, *i.e.*, relatively independent of the stage of differentiation of the cell in which it is contained or it can be inducible, *i.e.*, induced be specific environmental factors, such as the length of the day, the temperature, *etc.* or a promoter can be tissue-specific, *i.e.*, directing the expression of the coding sequence in cells of a certain tissue type. A "senescence" promoter is an inducible promoter that causes transcription to be initiated upon a certain event relating to age of the organism. A "heat shock promoter" is an inducible promoter that causes transcription to be initiated upon a change in temperature. An example of a heat shock protein promoter is the Soybean Gmhsp promoter. In addition to these inducible promoters, one of skill will realize that other inducible promoters can be used. For example, a wound induced promoter, like LAP. See, US Patent No. 5,962,670.

10

20

25

30

The term "purified" denotes that a polynucleotide or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the polynucleotide or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The term "specifically hybridizes" refers to a nucleic acid probe that hybridizes, duplexes or binds to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and

annealing conditions, see, for example, Sambrook et al., Molecular Cloning: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) ("Sambrook") or CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel et al., ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

5

10

15

20

25

30

The term "stringent conditions" in the context of polynucleotide hybridization experiments such as Southern and northern hybridizations refers to sequence dependent, binding and washing environments. An extensive guide to the hybridization of polynucleotides is found in Tijssen (1993) LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY-HYBRIDIZATION WITH NUCLEIC ACID PROBES part I chapter 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T<sub>m</sub>) for the specific sequence at a defined ionic strength and pH. The T<sub>m</sub> is the temperature (under defined lonic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T<sub>m</sub> for a particular probe. An example of stringent hybridization conditions for hybridization of complementary polynucleotides which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at between 40 and 50°C, preferably 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at from 70 to 80°C with 72°C being preferable for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at about 60 to 70°C, preferably 65°C for 15 minutes (see, Sambrook, supra for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 40 to 50°C, preferably 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6x SSC at 35 to 45°C, with 40°C being preferable, for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Polynucleotides which do

not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, e.g., when a copy of a polynucleotide is created using the maximum codon degeneracy permitted by the genetic code.

The term "transgenic plant" refers to a plant into which exogenous polynucleotides have been introduced and their progeny. Typically, cells of a plant are transformed with the exogenous polynucleotide and a transgenic plant is regenerated from the transformed cells. The regenerated plant is then bred to produce a strain of transgenic plants.

"Xylanase" (EC 3.2.1.8) refers to a well described class of gylcosyl hydrolases that hydrolize xylan. Commercial applications of xylanase include the degradation and bleaching of wood pulp for paper making. Xylanase can also be added to animal feed to improve the digestibility of plant matter. Typically, commercial xylanase is derived from fungi. A preferred xylanase is derived from *Trichoderma*.

#### Preferred Embodiments

5

10

15

20

25

30

Plant cell walls contain a range of alkali-labile ester-linked phenolic acids. In particular, grass cell walls are characterized by the presence of large amounts of esterified ferulic and p-coumaric acids (mainly in their E configurations), linked to arabinoxylans at the C5 of arabinose. These are released as ferulated oligosaccharides (FAX and PAX) by cellulase treatment but *in vivo* provide a substrate for peroxidase-catalyzed cross-linking of cell wall polysaccharides and lignin. The high levels of these phenolic acids and their dimers have a dramatic influence on the mechanical properties, digestibility and rates of digestion of grasses by ruminants.

Previous work has shown that ferulic acid is the predominant p-hydroxycinnamic acid esterified to grass polysaccharide but until recently the only ferulic acid dehydrodimer to have been isolated was 5,5'-diferulic acid. Recently new dehydrodiferulate dimers and cyclobutane-type dimer mixtures have been isolated from plant cell walls (Waldron, et al., Phytochemical Analysis 7:305 (1996)). As can be seen in Figure 1, these mixtures are present in large amounts in grass cells. Ether linked ferulic acid-coniferyl alcohol dimers, have also been isolated from cell walls (Jacquet, et al., Polyphenol Comm. Bordeaux

pp451 (1996)) establishing for the first time that ferulate esters are oxidatively copolymerized with lignin precursors which may anchor lignins to cell wall polysaccharides. The yield of these dimers in grass cells indicates that phenolic dehydrodimer cross-linking of cell wall polysaccharides is much more extensive than was previously thought.

An enzyme system has been reported from parsley endomembranes that catalyses the ferulation of endogenous polysaccharide acceptors from feruloyl CoA, pointing to the ER/golgi as the site of polysaccharide esterification and the CoA ester as the physiological co-substrate (Meyer, et al., FEBS Lett. 290:209 (1991)). Further evidence for this has been found in water-soluble extracellular polysaccharides excreted in large amounts into the medium by grass cell cultures. This material is highly esterified with ferulic and p-coumaric acid at levels similar to the cell walls of the cultured cells.

Feruloyl esterase activity has been detected in several fungal species including, anaerobic gut fungi, yeasts, actinomycetes, and a few fiber-degrading ruminal bacteria, which enables them to de-esterify arabinoxylans and pectins.

15

20

25

30

Two ferulic acid esterases (FAE), distinguished on the basis of molecular weight and substrate specificity, have been isolated from *Aspergillus niger* and have been shown to quantitatively hydrolyze ferulic acid and release dehydrodiferulate dimers from plant cell walls. Furthermore, FAE has been observed to act synergistically with xylanase to release ferulic acid from plant cell walls at a higher rate. Recently, a ferulic acid esterase (FAE) gene has been cloned from *Aspergillus niger* (Michelson, *et.al.* European Patent Application No. 9510370.1). The inventors have found the recombinant enzyme releases ferulic acid and diferulate dimers from grass cell walls in a concentration dependent manner and that this enzyme is stable at 30°C pH 5.0 in the presence of substrate and has a half life of 61 h at 30°C in the presence of vacuolar extracts (pH 4.6) of grass cells. This gene was, therefore, a candidate for targeted and indicible expression of FAE in grasses (*e.g.*, *Lolium multiflorum*).

The present invention provides for methods of changing the cell wall structure of transgenic plants and therefore, making them more digestible. The method comprises introducing a ferulic acid esterase coding sequence into the

cells of a plant. Operably linked to the coding sequence is a promoter that can be either constitutive or inducible and signal sequences that serve to target expression of the coding sequence in the desired organelle in the desired cell of the plant. The signal sequences can be either or both N terminal or C terminal sequences.

Optionally, a second and/or third coding sequence is introduced into the plant. It is preferred that a fungal xylanase coding sequence be coexpressed with the FAE coding sequence. .

This invention also provides for transgenic plants which contain FAE1 coding sequences, leading to more digestible grasses.

10

20

25

30

,

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions 15 involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. Basic texts disclosing the general methods of use in this invention include Sambrook, et al., MOLECULAR CLONING, A LABORATORY MANUAL, 2ND ED. (1989); Kriegler, GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL (1990); and Ausubel et al., (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1994)).

#### Α. **Isolation of Polynucleotides**

The isolation of the polynucleotides, e.g., FAE1 and xylanase of the invention may be accomplished by a number of techniques. See, for example, copending US application 08/952,445 which describes the isolation of a FAE from Aspergillus niger, and copending US application 09/658,772 which describes the isolation of a xylanase from T. reesel.

For instance, oligonucleotide probes based on the sequences cited here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g., using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be

packaged into the appropriate vector. To prepare a library of cDNA from a specific cell culture, e.g., Aspergillus niger, mRNA is isolated from the culture and a cDNA library containing the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based upon the sequence of a known polynucleotide such as the polynucleotides cited here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. In addition to probes derived from known polynucleotides, degenerate probes may be used. Techniques for making and using degenerate probes are well known in the art and can be found in Sambrook and Ausube).

10

20

30

Alternatively, the polynucleotides of interest can be amplified from polynucleotide samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone polynucleotides that code for proteins to be expressed, to make polynucleotides to use as probes for detecting the presence of the desired mRNA in samples, for polynucleotide sequencing, or for other purposes.

Appropriate primers and probes for identifying ferulic acid esterase-specific genes, as well as xylanase sequences, from fungi and plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San Diego (1990). Reaction components are typically: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μM dATP, 200 μM dCTP, 200 μM dGTP, 200 μM dTTP, 0.4 μM primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec., followed by 72°C for 5 min.

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers, et al., Cold Spring Harbor Symp. Quant. Biol. 47:411-418 (1982), and Adams, et al., J. Am. Chem.

Soc. 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Suitable sources for the ferulic acid esterase used in this invention include but are not limited to, *Neurospora crassa*, *Aspergillus spp.* and specifically, *Aspergillus niger*. The xylanase used in this invention can be derived from any suitable source including, but not limited to, *Trichoderma reesei* and *Aspergillus spp*.

10

15

20

25

30

5

#### B. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of plant species are well known and described in the technical and scientific literature. See, for example, Weising, et al., Ann. Rev. Genet. 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length FAE1 protein, will preferably be combined with transcriptional and translational initiation and targeting regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant under the desired conditions.

Promoters can be identified by analyzing the 5' sequences of a desired gene. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. Messing, *et al.*, in GENETIC ENGINEERING IN PLANTS, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, et al., Plant Cell 1:855-866 (1989); Bustos, et al., Plant Cell 1:839-854 (1989); Green, et al., EMBO J. 7:4035-4044 (1988); Meler, et al., Plant Cell 3:309-316 (1991); and Zhang, et al., Plant Physiology 110:1069-1079 (1996)).

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumafaciens*, the actin and ubiquitin promoters and other transcription initiation regions from various plant genes known to those of skill. A particularly preferred constitutive promoter is the rice actin promoter (see, McElroy, *Plant Cell*, 2:163 (1990)).

10

15

20

30

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots or seeds.

In one aspect of the instant invention, expression of FAE occurs after the the plant has been cut, removed from the ground or ingested. Thus an appropriate promoter would be a senescence promoter. For example, *BFN1* has recently been shown to be a nuclease expressed in senescing leaves, Perez-Amador, et al., Plant Physiol. 122:169 (2000). Similarly, SAG12, a cysteine protease is also found in senescing leaves (Noh & Amasino, Plant Mol. Biol. 41:181 (1999). In a preferred embodiment, the promoter from the *gem* gene of Festuca pratensis is used to direct expression of FAE in senescing leaves.

In another aspect, the FAE would be expressed upon Ingestion by a foraging animal. Exemplary promoters for this aspect would include Soybean Gmhsp 17.5 promoter and the leucine aminopeptidase (LAP) promoter. The

GMhsp promoter is from a heat shock protein gene and initiates expression if the temperature of the environment is increased. In the laboratory, an increase of 15°C for 2 hours is the preferred heat shock. However, in non-laboratory conditions suitable increases in temperature will occur in silos and in the rumen of animals that have ingested the plants of this invention. The LAP promoter initiates the expression of the FAE gene upon wounding of the plant. Such wounding would occur after cutting the plant or after mastication by a foraging animal. Tissue specific promoters that could be used in this invention include promoters of genes that are differentially expressed in the leaves of grasses. An example of a leaf specific promoter is the *rbcS* promoter of tomato (*Proc. Nat'l Acad. Sci. USA* 84:7104 (1987)). This promoter normally regulates a gene determined to be important in photosynthesis.

For proper polypeptide expression, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural fungal gene, from a variety of other fungal or plant genes, or from T-DNA. These sequences are well known and readily available to those of skill in the art.

15

20

25

30

In addition to a promoter and poly A+ sequences, the preferred expression vectors of this Invention also will contain signal sequences. These are polynucleotides found at the 5' and/or 3' ends of the coding region and serve to target expression of the gene to specific cellular organelles. These signal sequences can be both upstream or downstream of the coding region. Some preferred examples of upstream signal sequences include the barley aleuraln sequence (Rogers, *Proc. Nat'l Acad. Sci. USA* 82:6512 (1985) which targets vacuoles and the Aspergillus apoplast signal. This signal sequence targets expression to the apoplast.

In addition to targeting expression to specific organelles, it may be desireable to retain the expressed FAE in the Golgi or endoplasmic reticulum. The well known ER retention signal, KDEL, can be added to the 3' end of the coding polynucleotide.

The vector comprising the expression cassettes (e.g., promoters and/or coding regions) of the invention will typically comprise a marker gene which

confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to hygromycin, kanamycin, G418, bleomycin, or herbicide resistance, such as resistance to chlorosluforon or Basta.

5

10

15

20

25

30

#### C. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment or the constructs may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microInjection of plant cell protoplasts. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional Agrobacterium tumefaciens host vector. The virulence functions of the Agrobacterium tumefaciens host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

See Dalton et al. (Co-transformed, diploid Lolium perenne (Perennial Ryegrass), Lolium multiflorum (Italian Ryegrass) and Lolium temulentum (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports (1999) 18(9), 721-726) for exemplary methods for culturing and transformation of grasses.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, et al., Embo J. 3:2717-2722 (1984). Electroporation techniques are described in Fromm, et al., Proc. Natl. Acad. Sci. USA 82:5824 (1985).

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, et al., Science 233:496-498 (1984), and Fraley, et al., Proc. Natl. Acad. Sci. USA 80:4803 (1983). US Patent 5,591,616 discloses Agrobacterium mediated transformation techniques in monocotyledons.

Ballistic transformation techniques are described in Klein, et al., Nature

**327**:70-73 (1987). In a preferred embodiment, a particle in-flow gun (PIG) is used to transform the plant cells of this invention.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as improved digestibility. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a blocke and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, et al., Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176, Macmillian Publishing Company, New York, 1983; and Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, et al., Ann. Rev. of Plant Phys. 38:467-486 (1987).

To determine the presence of or increase of FAE1 activity, an enzymatic assay can be used or an assay to measure increases and decreases in rates of fermentation. These assays are readily available in the literature and those of skill in the art can readily find them.

15

20

25

30

One of skill will recognize that other assays can be used to detect the presence or absence of FAE1. These assays include but are not limited to; immunoassays and electrophoretic detection assays (either with staining or western blotting).

The polynucleotides of the invention can be used to confer desired traits on essentially any plant. However, the main utility of this invention is in the improved digestibility of forage plants. Thus, it is envisioned the transgenic plants of this invention will include but not be limited to the following genera. Lollum, Festuca, Triticum, Avena, and Medicago. The FAE1 genes of the invention are particularly useful in the production of transgenic plants in the genus Lolium.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be

introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

As mentioned above, the transgenic plants of this invention can be used as a foraging crop for animals, such as cattle, sheep, goats and horses. In addition, the methods of this invention can be used to transform any plant into which FAE expression is desired. For example, it is advantageous to break down cell walls during biomass conversion or during processing of plants for foodstuffs. This invention would help to achieve this goal more effectively and inexpensively.

10

15

20

30

1.52

The inventive methods herein may also be used to provide additional enzymes to enhance the availability of fermentable sugars in plants. Plant carbohydrates may be subject to further modification, either exogenously or endogenously, by the action of other enzymes. Such enzymes include, but are not limited to, endoglucanases, xylosidases and/or cellbiohydrolases. These enzymes may be provided either in an expression cassette provided for herein (i.e., endogenous) or applied to the plant cell walls (i.e., exogenous) to enhance the availability of mono- and/or di-saccharides.

Plants other than grasses may find a use in the present Invention. For example, corn (or maize) is specifically contemplated to be useful. The grass Festuca is similar to maize in cell wall structure and therefore provides a good model of the ability to enhance fermentable carbohydrates in corn. Other useful plants contemplated for use in the present invention are Festuca, Lolium, Zea, Avena, Sorghum, Millet (tropical cereals), Miscanthus (a grass with potential for use as a biomass energy crop), Cenchrus, Dichanthium, Brachlaria and Paspalum (apomictic tropical range grasses) and Poa (Kentucky bluegrass).

Cell walls of forage grasses makes up 30-80% of forage dray matter representing a major source of energy for ruminants, but less than 50% of this fraction is digested by the animal. Conversion of low-value biomass to sugars and ethanol is also less than optimal due to the carbohydrate unavailability of the

feedstocks, including but not limited to bagasse, race straw, com stover and corn fiber.

Ferulic and other hydroxycinnamic acids are ester linked to arabinosyl residues in arabinoxylans, and play a key role in crosslinking xylans to liginin, resulting in less degradable cell walls. Ferulic acid esterase (FAE) can release both monomeric and dimeric ferulic acid (FA) from arabinoxylans making the cell wall more susceptible to further enzymatic attack. Transgenic plants have been produced expressing an FAE gene following microprojectile bombardment of cell cultures. Measurements of the level of FAE activity from different vectors targeting FAE to the vacuole, ER and apoplast under constitutive or inducible (heat shock) promoters shows that at least for constitutive expression of vacuolar targeted FAE, the activity was highest in young leaves and increased along the leaf lamina. We also show that FAE expression results in release of monomeric and dimeric FA from cell walls on cell death and this was enhanced several fold by the addition of xylanase. An effect of FAE expression on the monomeric and dimeric cell wall ester linked ferulate content in comparison to control (nontransformed) plants is seen. Generally, the lower the levels of monomers and, in particular, dimers of hydroxycinnamic acids in leaves, the higher the digestibility and/or availability of complex carbohydrates for conversion.

Senescence is the terminal phase in leaf development and occurs without grouth or morphogenesis. Therefore the metabolism/physiology of this stage of the leaf's lifespan can be targeted directly for alteration with minimal detrimental impact on early development. Senescence follows leaf maturity and is associated with the expression of specific genes. These genes and their controlling elements can be exploited to manipulate development, adaptation, productivity and quality traits in crop plants. There seems to be good conservation of senescence physiology across the range of higher plant species and thus these promoters are useful in the present invention.

20

30

The following preparations and examples are given to enable those skilled in the art to more clearly understand and practice the present invention. They should not be considered as limiting the scope and/or spirit of the invention, but

merely as being illustrative and representative thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); µM (micromolar); N (Normal); mol (moles); mmol (millimoles); µmol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); kg (kilograms); µg (micrograms); L (liters); ml (milliliters); µl (microliters); cm (centimeters); mm (millimeters); µm (micrometers); nm (nanometers); ° C. (degrees Centigrade); h (hours); mln (minutes); sec (seconds); msec (milliseconds); Ci (Curies) mCi (milliCuries); µCi (microCuries); TLC (thin layer achromatography); Et (ethyl), Me (methyl).

10

20

25

## Example 1

Preparation of Enzyme Encoding DNA Sequences

A genomic clone for FAE1 (see Figures 1-3) was used as the starting point for the preparation of an intronless FAE1 encoding DNA sequence. The sequence for the genomic clone is given in Figures 2 and 3. Separate fragments for both FAE exons were recovered by PCR from a 5.5kb EcoRI fragment of the genomic clone in pLITMUS28, and 'cDNA' created by overlapping PCR. See Figure 4.

Two 5' primers were used. FAE-S5 which amplifies the entire reading frame (including the Aspergillus signal), and FAE-N5 which amplifies only the mature protein (i.e. has no signal). A number of codons are optimised (underlined in primer sequences below). The overlap product may be derived from either FAE-I5 (wild type) or FAE-I3 (conserved Ser changed to Ala) primers, allowing production of enzymatically inactive protein to check toxicity. As shown in Figure 5, overlapping of PCR products made with FAE-I5 and FAE-I3 creates two possible uninterrupted reading frames. If the complement to FAE-I5 serves as the template when recombined then the encoded protein retains the serine moiety and the esterase is functional (highlighted serine is at active site). If the FAE-I3 primer serves as the template the serine is replaced with an alanine and the esterase is inactivated (highlighted alanine in bottom amino acid sequence given in Figure 5).

Where possible, codon usage has been optimised in constructed reading

frames (codon choice based on published barley preferences).

FAE-I5 (SEQ ID NO: \_\_\_)

GGCGCCGAGGGAGTGGCCGGTCACGGTCAGCGCGTAGTCC 40-mer
FAE-I3 (SEQ ID NO: \_\_\_)

CCGGCCACGCCTCGGCGCCTCCCTGGCGGCACTC 35-mer
FAE-N5 (SEQ ID NO: \_\_\_)

CTAAAGCTTACCATGGCGGCCGCCTCCACGCAGGGCATCTCCGA 44-mer

FAE-S5 (SEQ ID NO: \_\_\_)

CTAAAGCTTAACATGAAGCAGTTCTCCGCCAA 32-mer
FAE-3 (SEQ ID NO: \_\_\_)

TCTAAGCTTGCGGCCGCCGCGCCAGGTGCATGCGCCGCTCATCCC

Example 2

Preparation of Vectors

Vectors had the general structure shown in Figure 6.

A. Plant transformation vector series

20

25

30

15

50-MER

Initial expression vectors were based on pCOR105 [rice actin promoter - McElroy et al. MGG 231:150-160 (1991)] (Figure 7). pCOR105 Not and SstII sites were first destroyed [cut with NotI and SstI, followed by heat inactivation and T4 DNA polymerase treatment in the presence of dNTPs] using standard methods as described in Maniatis et al. or following the manufacturer's instructions for enzymes to simplify subsequent Not cassette manipulation and allow use of unique Sst site (see below).

The nos terminator from pMA406 (Ainley & Key (1990) PMB 14:949-60) was amplified by PCR using primers TER5 and TER3 to generate a fragment with the following sequence (SEQ ID NO:\_\_\_):

(Pst1) (Not 1)

(AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG

AAGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT

GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT

AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTTATGATTA

GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC

AAACTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATA

AGCTT CTA GATCT (where K=G or T)

10 (HindIII) (Xbal)

15

20

25

30

A redundancy in the TER5 primer (GCGKAA) creates fragments having either a stop codon (TAA) or glutamate codon (GAA) in one reading frame. The glutamate codon is in frame with a downstream KDEL motif.

The fragment and modified pCOR105 vector were cut with Pstl and Xbal, according to manufacturers instructions, relevant fragments gel-purified, ligated with T4 DNA ligase and transformed into *E. coli*. Resulting clones were then sequenced to establish which TER5 alternatives were present.

Initial FAE expression vectors were then constructed from these vectors by inserting FAE-S5/FAE-3 PCR products (T4 DNA polymerase 'polished' in the presence of dNTPs, purified and digested with Notl, cloned into EcoRV and Notl digested vector) or FAE-N5/FAE-3 PCR products (purified and Notl digested, cloned into Notl digested and calf intestinal alkaline phosphatase treated vector).

The Initial pCOR105-nos terminator clones were also modified by the addition of ALE-5/ALE-3 PCR products (encoding wild-type and modified barley aleurain signal peptides, see below for details). The products were 'polished' with T4 DNA polymerase in the presence of dNTPs, purified and cut with Notl, then cloned into EcoRV and Notl digested vectors. Addition of the ALE sequences creates a series of vectors which can express a reading frame inserted at the Notl or Ncol sites as a fusion to the barley aleurain signal, with or without vacuolar targetting motif, and with or without an ER retention motif. Hind III sites flanking the translation initiation codon and transcriptional terminator allow easy

movement of transcription units between expression vectors providing different promoter sequences. (See Figure 8 depicting the generic ALE-TER vector.)

Vector sequences were confirmed by sequencing. Two artifacts were found. Firstly, the redundant codon in TER5 was found to be AAA in one clone, which was subsequently used as the source of all KDEL fusions (le peptide sequence is KPLKDEL, rather than EPLKDEL as designed). See Figure 9. Secondly, an additional base is found at the site of the redundant codon in one clone, creating a frameshifted terminal peptide (ETTEG, Figure 10) which was used as a control in some constructs.

10

15

20

25

30

Exploitation of the modular arrangement of signal peptides in the above vector series allowed various combinations of FAE and targeting motifs to be created using standard molecular biology procedures (i.e., restriction digest, purification of relevant fragments and ligation as appropriate). For example, the Notl fragment containing the FAE reading frame was inserted into the Notl site of the frameshifted clone described above to create vector pTP3.1. The native Aspergillus COOH-terminus was inserted into a FAE-S5/FAE-3 clone as a Sphl (T4 DNA polymerase polished) – Ncol fragment from the FAE genomic clone (replacing the Notl (T4 DNA polymerase polished) – Ncol fragment), creating vector pTP4a2, which then encodes the entire, unmodified, Aspergillus FAE. Replacement of the Sall/Xbal fragment of pTP3.1 with that of pTP4a2 then created pTP11.1, which encodes FAE with a native Aspergillus COOH-terminus but a barley aleurain N-terminal signal.

Briefly, other vectors made in this series were; pTP8.5, the FAE Notl fragment inserted into the Notl site of an ALE-frameshifted COOH-terminus construct, aleurain N-terminus; pTP5.1, replacement of the native Aspergillus COOH terminus with a KDEL peptide (Notl/Xbal fragment exchange), Aspergillus N-terminal signal retained; pTU4.4, BamHI fragment of pTP11.1 replaces BamHI fragment of pTP5.1, creates FAE reading frame fused to heterologous N- and C-termini (aleurain signal and KDEL).

Vectors in which the aleurain vacuolar targeting motif NPIR was replaced by NPGR (found to be inactive in some plant assays) were created by replacing an EcoRV/NotI fragment with ALE PCR product which had been cut with Acci

(T4 DNA polymerase polished) and Notl (vectors pTT5.5 and pTT5.14, Aspergillus COOH-terminus). The BamHl fragment of pTT5.5 was used to replace that of pTP5.1 to produce pTU5, creating an FAE reading frame fused to heterologous N- and C-termini (NPGR modification of aleurain signal and KDEL).

The aleuraln signal was also modified by PCR mutagenesis to remove the vacuolar targeting NPIR motif in its entirety (directed by primer ALECUT, which contains a NotI site to allow exchange of BgIII/NotI fragments). NPIR deletion was created in this way in pTP11.1 (creating pUA4.4), and in pTP5.1 by exchange of BamHI fragments with pUA4.4 (creating pUG4).

Finally, PCR mutagenesis, using overlap of fragments generated by primers GLY3 and GLYB, was also used to alter a potential glycosylation site (asparagine codon changed to aspartate, as carried out for example in Chen, H.M., C. Ford & P. J. Reilly (1994) Biochem J 301 275-281 Substitution of asparagine residues in Aspergillus awamori glucoamylase by site-directed mutagenesis to eliminate N-glycosylation and inactivation by deamidation; see sequence data for exact change, vector pTP10.1).

PCR primers
TER-5 (SEQ ID NO:\_\_)
AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
AAGCAGATCGTTCAAACATTTG 72-MER (The KDEL stop codon is underlined.)
TER-NOT (SEQ ID NO:\_\_)
AAGACTGCAGACCATGGCGG 20-MER
TER-3 (SEQ ID NO:\_\_)
AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC
ALECUT (SEQ ID NO:\_\_)
CTAGGCGGCCGCGGGGAGGAGGCGACGGCGAC
GLYB (SEQ ID NO:\_\_)
GAGGGTGTATTCGGTATCGAGTTGCAGGTTCGTATC
GLY3 (SEQ ID NO:\_\_)
CTCGATACCCATTACACCCTCACGCCTTTCGA

35 B. Construction of different promoter vectors
Various promoters were used to optimize expression and to establish
constitutive, heat-shock inducibility and senescence enhancement.

i. Rice actin promoter and 1st intron

10

15

Initial vectors (Figures 11 and 12) were constructed from pCOR105 which was subsequently found to contain a 5bp deletion relative to the published sequence which destroys the Accl site (GTAGGTAGAC, deleted bases underlined) and may affect splicing at the adjacent 3' site. The original rice actin sequence in this region (GTAGGTAG) was therefore restored using oligonucleotide NCO-ACT (CTCACCATGGTAAGCTTCTACC TACAAAAAAGCTCCGCA) by replacing the Bgill/HindIII fragment with a PCR product, to produce vector pPQ10.1.

A rice repetitive element is present in the upstream region of the actin promoter used in pCOR105; as this may have unpredictable effects on vector expression it was removed from pPQ10.1 by deletion of the Kpnl/EcoRl fragment (end-filled with T4 polymerase and ligated following digest, restoring EcoRl but not Kpnl), to produce vector pGT6. The Hindill fragment containing the FAE reading frame and nos terminator of pTP3.1 (see Example 2A) was then inserted into pGT6 to produce construct pJO6.3.

#### ii. Sovbean heat-shock promoter

A soybean heat shock promoter from a 23kD HSP was obtained from pMA406 (Ainley & Key (1990) PMB 14:949-60). This promoter when fused to  $\beta$ -glucuronidase (Jefferson et al 1987 EMBO J 6:3901-3907) had previously been shown to be inducible by a 10°C heat-shock and show stable expression for 24-48 hours (data not shown).  $\beta$ -glucuronidase fusions are a sensitive and versatile fusion marker in higher plants. The construction of the co-integration HS vectors is given below.

25

30

15

20

iii. Senescence enhanced expression (See1) promoter from Lolium multiflorum

The promoter and signal sequence (including NPIR motif) of the LSee1 gene was amplified from *Lolium multiflorum* cv Tribune with ollgonucleotides SEE-NCO and SEE-VAC, and cloned as an Asp718/Noti replacement of the promoter region of vector **pTP11.1**. Following sequencing to screen for PCR artifacts, one of three identical clones was chosen (**pUB8.11**).

The See1 promoter from maize has been doned previously and has

EMBL accession number is AX050343. See WO0070061.

The Lolium version of See1 was also cloned previously (Qiang Li (2000) Studies on leaf senescence and its genetic manipulation in *Lolium mutiflorum* PhD Thesis University of Wales, Aberystwyth) and has been shown to be senescence inducible when used to drive both β-glucuronidase and the Agrobacterium ipt gene.

An apoplast-targeted derivative was constructed by amplifying the Potato Protease Inhibitor (PPI) motif with primers PPI-AP6 and SEE-ATG, and cloning the product as an NgoMIV/NotI fragment into pUB8.11 (NgoMIV partial digest), to produce vector **pJQ5.2**. This vector has both the senescence induced promoter and the apoplast target sequence with the gene to be expressed inserted downstream of the apoplast sequence.

	PCR Primers
	SEE-VAC (SEQ ID NO:)
15	AACCATGGCGGCCGCGCGCTCGGTGACGGCCGGAT
	SEE-NCO (SEQ ID NO:)
	TTCGGTACCATGGCCAGGTATAATTATGG
	SEE-ATG (SEQ ID NO:)
	CTGCGCCGGCGAGATGGMCGTGCACAAGGAG

20

25

30

#### C. Construction of targeting sequences

In order to examine whether or not the localization of the enzyme would have an effect on the phenolic acid content of the cell wall various signal sequences were utilized. The targeting sequences were added either to the N-terminus or to the C-terminus of the gene of interest.

- i. N-terminal signal sequencesSix N-terminal signal sequences were utilized:
  - (a) The native Aspergillus end of FAE, plus excretion signal [apoplast localisation]

This is from the original clone and has the peptide sequence:

MKQFSAKHVLAVVVTAGHALAASTQGI.

(b) The mature Aspergillus end, with no excretion signal [cytoplasmic localisation]

Peptide sequence is MAAASTQGI (underlined motif is common to all constructs). Truncation of the signal sequence in (a) above was carried out by PCR with mutagenic primer FAE-N5.

5

10

15

(c) The barley aleurain signal, including intact NPIR motif [vacuole localisation]

The barley aleurain vacuolar signal sequence (See Figure 13; Swissprot database accession number P05167) was derived entirely from overlapping primers (ALE-5, ALE-3, ALE-CUT ALE-CAP-5 and ALE CAP-3). Following primer annealing at 37°C and extension with T4 DNA polymerase in the presence of dNTPs according to manufacturers instructions, PCR with flanking primers ALE-5 and ALE-3 was carried out. The product was 'polished' with T4 DNA polymerase, purified, digested with Notl and cloned into EcoRV/Notl digested pCOR105-nos terminator vector (see above). ALE-3 contains redundancies so that clones encoding NPIR or NPGR motifs may be recovered. Two versions of the signal, with and without the vacuole targeting motif, were produced, to give putative vacuolar NPIR and apoplast (NPGR) signal sequences.

	PCR Primers
	ALE-5 (SEQ ID NO:)
20	GGAATTCGTAGACAAGCTTACMATGGCCCACGCCCGCGTCCT 41-MER
	ALE-3 (SEQ ID NO:)
	TATCCATGGCGGCCGCGGGTCGGTGACGGGCCGGMYCGGGTTGGAGTC
	GGCGAA 55-MER
	ALE-CUT (SEQ ID NO:)
25	CTAGGCGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGG
	ALECAP-5 (SEQ ID NO:)
	GCGACGCCGACGCCGTGGCCAGCACGCCGAGCGCCAGGAGGAGG
	ACGCGG 54-MER
	ALECAP-3 (SEQ ID NO:)
0	TCGCCGTCGCCTCCTCCTCCTCGCCGACT 33-MER

(d) The barley aleurain signal, mutated to a NPGR motif [cytoplasmic localisation]

(e) The rat sialyl transferase golgi targeting motif [golgi localisation]

A Golgi targeting vector, pJQ3.2, was made by Inserting a reading frame encoding the relevant rat sialyl transferase (RST) motif (See Figure 14. RST motif shown to function in plants by Boevink P, Oparka K, Cruz SS, Martin B, Betteridge A, Hawes C, (1998) PLANT JOURNAL 15 441-447 Stacks on tracks: the plant Golgi apparatus traffics on an actin/ER network) into vector pPQ10.1, and replacing the EcoRl/Notl promoter/signal fragment of pJ06.3 with the fragment from this vector. Briefly, the RST motif was constructed by annealing oligonucleotides RST-F1A, RST-F1B, RST-F2A and RST-F2B, and amplifying the product with RST-5AD and RST-3A. This product was cloned and sequenced. Clones were found to have a deletion which was corrected by PCR with RST-RPT, followed by overlap-PCR and cloning of products.

	PCR primers
	RST-5AD (SEQ ID NO:)
	ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA
20	RST-F1A (SEQ ID NO:)
	TTCCATGATCCACCCAACCTCAAAAAGAAGTTCTCCCTCTTCAT
	RST-F1B (SEQ ID NO:)
	AGAGTGATCACGGCGAAGAGGGGAGGAGACGAGGGAGAGACTTCT
	TIT
25	RST-F2A (SEQ ID NO:)
	TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCCTCCAA
	GCCAAGGA
	RST-F2B (SEQ ID NO:)
	CATTTGGAACTCCTTGGCTTGGAGGGTG
30	RST-3A (SEQ ID NO:)
	AACCATGGCGCCCCATTTGGAACTCCTTGGCT
	RST-RPT (SEQ ID NO:)

## TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCCTCC AAGCCAAGGA

(f) otif [cytoplasmic localisation]

5

30

(g) The potato protease inhibitor II (PPI) apoplast motif [apoplast localisation]

An apoplast targeting reading frame was designed to encode the relevant potato protease inhibitor II (PPI) motif (See Figure 15) and cloned into **pJO6.3**, to produce vector **pJQ4.9**. Briefly, the PPI motif was constructed by annealing oligonucleotides PPI-AP1, PPI-AP2, PPI-AP3, PPI-AP4, PPI-AP5 and PPI-AP6, and cloning this product as a HindIII/NotI fragment into vector pPQ10.1; the EcoRI/NotI promoter/signal fragment of pJO6.3 was then replaced with the equivalent fragment from the modified pPQ10.1 vector.

	PCR primers
15	PPI-AP1 (SEQ ID NO:)
	GGAATTCGTAGACAAGCTTACMATGGMCGTGCACAAGGAGGT
	PPI-AP2 (SEQ ID NO:)
	GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC
	PPI-AP3 (SEQ ID NO:)
20	CCTACCTCCTGATCGTSCTCGGCCTCCTCTTGCTCGT
	PPI-AP4 (SEQ ID NO:)
	CCTTGGCGTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG
	PPI-AP5 (SEQ ID NO:)
	GTGGACGCCAAGGCCTGCACCCKCGAGTGCGGCAACCTC
25	PPI-AP6 (SEQ ID NO:)
	GGAATTCGCGGCCGCCGGGCAGATGCCGAAGCCGAGGTTGCCGCACT

ii. C-terminal end signal sequences

Four C-terminal signal sequences were utilized:

(a) Native Aspergillus end, [CTW] (vacuole and apoplast vectors)

This was derived directly from the genomic clone (see Example 1) as a Nco1-Sph1 fragment (Sph end filled with T4 polymerase) which replaces the

Nco1-Not1 region of a standard actin -FAE vector (Not1 end filled with T4 DNA polymerase).

(b) Expression vector linker alone [CTW-PVAAA] (plant optimised Cterminus for vacuole, golgi and apoplast vectors)

CTW is the peptide sequence of the Aspergillus FAE COOH end and is here provided by oligo FAE3. In this primer the reading frame is extended to provide the additional amino acids PVAAA which are partially encoded by the Not1 site used for cloning downstream signals see c) and d) below. Some COOH amino acids /motifs may affect compartment targeting, the PVAAA sequences are expected to be neutral in this respect while the native Aspergillus end may not be.

(c) Linker plus KPLKDEL [first K is primer artifact, intended to be E] {ER retention vectors}

These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone. KDEL targeting has been demonstrated in plants by Denecke et al. ((1992) EMBO J 11: 2345-2355 Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope).

(d) Linker plus ETTEG [frameshift of (c)] (loss of ER retention - vacuole vectors)

These sequences are provided by primer TER5 Introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone (see Example 2A).

The KDEL signal is for ER retention, while others provide controls. A frameshift in the TER5 region [additional A] was used in subsequent constructs to destroy the ER KDEL retention signal.

The linker used in the above C-terminal targeting sequences was PVAAA.

D. Co-integration and co-transformation vectors.

Co-transformation vectors

5

10

15

20

25

30

A Hygromycin resistance gene driven by a CaMV345S promoter (pRob5) (35S-HYG-CMV in pUC18 (modified HYG, derived from pGL2) Bilang et al (1991)

Gene 100:247-50) was used for co-transformation experiments with pTT3 and pTP3.1, pJQ4.9, pJQ3.2, pJQ6.3, pJQ5.2, pUB8.1 1 vectors.

#### 5 Co-integration vectors

30

 Actin promoter constructs - pTR2.22, pTR6.1, pTR8.1, pTR9.4, pTR7.1, pTT5.5 and 5.1.

The CAMV35S-hyg region from pAJEB64TCA [a plant expression vector constructed by Andy Bettany at IGER containing CaMV-HYG from pTRA151 (Zheng et al 1991 Plant Physiol 97:832-835) (CaMV35S-HYG-tml terminator as clonable cassette in pUC4) cloned Into Kpnl site of pCOR105] was added as a HindIII fragment at the Kpnl site (T4 polymerase blunt) of pTP4a2, in divergent orientation to FAE to create pTR2.22. The FAE/Nos HindIII fragment of this vector was replaced as follows in co-expression vectors. From pTP5.1 for pTR6.1, from pTP10.1 to pTR8.1, from pTP11.1 to pTR9.4. Signal sequences of FAE in pTR2.22 were replaced as HindIII/BgIII fragments in pTR7.1 (fragment from pT09.1). PCR products (ALE5/ALE-G) was digested with Acc1 and T4 polymerase, polished, followed by Not1 digest and cloning into EcoRV/ Not1 digested pTR2.22 to give clones pTT5.5 and 5.1.

PCR primer
ALE-G
TATCCATGGCGGCCGCGGGTCGGTGACGGCCCGGGTTGGAGTC
GGCGAA

2. Actin promoter constructs -pUF1, pUA1K3, pUH4, pUH5, pUH6, pUH7, pUH8, pUH9.

The HygR gene from pAJEB64TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled Xbal site of pTP3.1, to give pHOX3. For ease of cloning the downstream HindIII site was destroyed to create pUA1K3 and replacement of the FAE/Nos terminator HindIII fragment in this vector was carried out as follows. From pTP5.1 for pUF1, from pTP11.1 for pUH4, from pTP8.5 for UH5, from pTT5 for pUH6, from pUA4.4 for pUH7, from pTU5 for pUH8 and from pUG4 for pUH9.

#### 3. Heat-shock promoter constructs - pUH10, pUH12, pUC5.11.

5

10

15

20

25

30

A co-transformation vector in which FAE is expressed from the soybean heat shock promoter was made by first modifying pMA406 to remove the nos terminator (BgIII linearised and gel purified, KpnI digested, T4 DNA polymerase polished in the presence of dNTPs and recircularised), and then inserting the FAE HindIII fragment from pTP11.1, creating pTT3.1, which encodes the full aleurain signal and the native Aspergillus COOH-terminus.

Following assays of various constructs, co-Integration vectors were constructed with FAE and HygR genes arranged in tandem.

The HygR gene from pAJEB-64-TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1, to give pHOX3 and subsequently excised as a HindIII/SacI fragment (partial SacI digest, relevant sites found in flanking pTP3.1 sequences) which was cloned into the HindIII/SacI sites of pMA406, in tandem orientation (vector pUH1a20). FAE sequences were then cloned into the HindIII site of pUH1a20 downstream of the heat -shock promoter (HindIII fragment from pTU5 for pUH10, HindIII fragment from pTT5 for pUH12). A pTP3.1 derivative was made by cloning the CaMV/HygR HindIII cassette from pAJEB-64-TCA in tandem orientation downstream of the FAE gene in pTP3.1, Inactivating the middle HindIII site by partial digestion and end-filling, and excising the combined FAE/HygR cassette as a single HindIII fragment, which was inserted at the HindIII site in pMA406 to produce pUC5.11.

#### Example 3

Transformation of Plant Cells

Eight to ten weeks old embryogenic *F. arundinacea* and *L. multiflorum* suspension cultures were bombarded either with a single co-integration plasmid DNA vector containing FAE and hyg resistance genes, or with a co-transformation vector containing FAE and with plasmid pROB5 conferring hygromycln resistance (CAMV35S-hpt- nos) using a Particle Inflow Gun (PIG) (Finer et al. (1992) Development of the particle inflow gun for DNA delivery to

plant cells Plant Cell Reports 11:323-328) and 1.5-3.0 µm gold particles as in Dalton et al (Dalton et al. (1999) Co-transformed diploid Lollum perenne (Perennial ryegrass), Lolium multiflorum (Italian ryegrass) and Lolium temulentum (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports. 18: 721-726) and Kuai et al (Regeneration of fertile transgenic tall fescue (Festuca arundinacea) plants with a stable highly expressed foreign gene. Plant Cell Tissue and Organ Culture (1999) 58:149-154). Transformants were selected with hygromycin (25 to 50mg /l) over a 10-12 week selection period at 25°C under continuous white fluorescent light (60 µE m<sup>2</sup> s<sup>-1</sup>) and plants regenerated via somatic embryogenesis as in Dalton et al 1999, supra. Regenerated plants were screened for FAE activity on transfer to soil and expressing plants grown to maturity in a containment growth room at 18°C under 16h fluorescent lights (350 μE m<sup>2</sup> s<sup>-1</sup>). Mature plants (6-8 weeks old) were re-assayed for FAE activity and fresh tissue harvested for Southern, Northern and Western analysis, and for self digestion analysis. The remaining tissue was freeze dried and powdered for cell wall structure analysis, in vitro-dry matter digestibility (IVDMD) determinations and for in-vitro gas production determinations of rates of tissue digestion.

10

15

20

25

30

#### Example 4

**Targeting of Expression Product** 

To verify that the targeting sequences are effective in delivering the gene the targeting sequences were operably linked to a green fluorescent protein GFP. The vector constructs are shown in Figure 16. Cells were transformed by particle bombardment as in Example 3. Localization of the GFP could be visualized under a microscope 1 day after bombardment (i.e., shooting). See Figure 16.

### Example 5 FAE1 activity

Plants regenerated from transformed cells showed FAE activity in all plant tissues tested. Cells were transformed as above under the direction of the ER and APO targeting sequences. FAE activity in transformed Festuca arundinacea

leaves of different ages was elevated compared to control (untransformed) plants. See Figures 17 and 18.

Similar results were seen with Lolium mutiflorum leaves at different ages transformed as above under the direction of vacuolar, ER and APO targeting sequence. See Figures 19 and 20.

FAE expression under a heat shock promoter can also be induced. (Data not shown.)

Thus, we have demonstrated FAE expression in Festuca and Lollum leaves under constitutive and HS promoters with effective FAE targeting to the vac, ER and apo.

FAE assay

5

10

25

FAE activity was determined in soluble extracts of fresh (or frozen at -70oC) leaves or cell cultures (0.5g) with 0.1M NaAc, pH 5.0 extraction buffer. Extracts were incubated with 24mM EF (ethyl 4-hydroxy-3-methoxycinnamate) or 1% FAXX as substrate, at 28°C for 24hrs and FAE activity calculated as the amount of ferulic acid released. FAE activity was also determined by measuring the release of monomeric and dimeric ferulic acid from self-digested leaf or cell culture samples. Fresh, or frozen, leaves or cell cultures (0.5g) were ground in 0.1M NaAc, pH5.0 extraction buffer in the presence and absence of xylanase (1000U GC140/sample) without added substrate and incubated at 28°C for 72hrs. Following incubation, and centrifugation, soluble extracts were loaded onto an activated reverse phase C18 μNova sep-pak column (Waters), eluted with 100% MeOH and the MeOH sample analysed by HPLC.

Example 6

Chemical Analysis of Cell Wall Extracts
Ester bound compounds were extracted from freeze dried powdered leaves
or cell cultures (50 -100mg) with NaOH (5ml of 1M) followed by incubation at 25°C
for 23hrs under N2. After centrifugation and acidification of the soluble extract with
concentrated HCl, the extracted phenolics were loaded onto an activated reverse
phase C18 µNova sep-pak column (Waters) and eluted with 100% MeOH. and the
MeOH sample analysed by HPLC.

HPLC was carried out with methanol: 5% acetic acid either with a 35-65% MeOH gradient in 15min (FAE assay) or with a 30-70% MeOH gradient in 25 min (monomer and dimer cell wall components) at 2ml/min on a µNova Pak C18 8x10 RCM (Waters). Extracts were detected and quantified with a diode array detector (240-400nm Waters 996PDA) monitored at 280nm for aldehydes and 340nm for hydroxycinnamic acids.

Levels of esterified monomeric and dimeric hydroxycinnamic acids in Festuca arundinacea plants expressing FAE under VAC, and ER and APO targeting sequences are reduced compared to control (untransformed) plants. The results can be seen in Figure 21 and 22, respectively. Thus, we show where this does not result in reduced cell wall phenolics in growing plants with vac targeting but does result in lower phenolics with ER and apo targeting. In addition.

Levels of esterified monomeric and dimeric hydroxycinnamic acids in Festuca arundinacea plants expressing FAE are not significantly reduced when FAE is VAC, targeting (Fig 21) which is as predicted for correct vacuolar targeting, but are significantly reduced, as predicted, in some plants when FAE was ER and APO targeted, compared to control (untransformed) plants. The results can be seen in Figure 22.

20

25

30

15

#### Example 7

In vitro dry matter digestibility. (IVDMD)

The *in vitro* dry matter digestibility (IVDMD) was estimated on 1.0 g dry weight of leaf or cell culture tissue using the pepsin/cellulase method of Jones and Hayward (The effect of pepsin treatment of herbage on the prediction of dry matter digestibility from solubility in fungal cellulase solutions. Journal of the Science of Food and Agriculture (1975) 26:711-718).

We show that the presence of FAE in the plants results in higher digestibility of the leaves. This may be due to internal FAE activity acting on normal cell walls with vacuole located FAE and to both FAE activity and the lower cell wall crosslinking with ER and apo targeted FAE (as also found with cell cultures).

End point digestibility as determined by IVDMD were higher in leaf tissue of some transformed plants of Festuca expressing FAE, compared to control (untransformed) plants. Examples are shown where vacuolar, ER or apoplast targeted FAE under a constitutive actin promoter have been effective at increasing IVDMD. Similar results were obtained with in leaves of Lolium, but were less pronounced.

The results can be seen in Figures 23 and 24.

10

15

20

30

#### Example 8

In vitro gas production measurements

In each experiment, 1.0-g samples of freeze dried powdered leaf tissue or cell culture were fermented in three 165-ml capacity serum bottles according to the pressure transducer technique of Theodorou et al. (Theodorou et al. (1994) A new gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. Animal Feed Science and Technology 48: 185-197). Grab samples of rumen-digesta were taken at 8.00 h before the moming feeding from fistulated wethers fed grass hay, and transported to the laboratory in a prewarmed (39°C) vacuum flask. The microbial inoculum and culture media were prepared as described by Theodorou et al. (1994). Each serum bottle received 10 ml of microbial inoculum, 85 ml of buffer and 4 ml of reducing agent.

At the end of the incubation period, (144h) the contents of each serum bottle were filtered through pre-weighed sintered glass funnels and freeze dried to constant weight. Dry matter loss was calculated as the difference between the dry weight of the sample pre- and post-incubation. Additionally, the concentration of volatile fatty acids (VFA) in the liquid fraction of the culture media at the end of the 144-h incubation period was determined by gas chromatography. A Chrompack CP 9000 chromatograph fitted with an automatic sampler (Chrompack 911) and a flame-ionisation detector, linked to a Dell PC with A1-450 integration software, was used for VFA quantification.

Gas production data were fitted to the model of France et al. (France, J., Dhanoa, M.S., Theodorou, M.K, Lister, S.J., Davies. D.R. and Isac, D. 1993. A model to interpret gas accumulation profiles associated with *in vitro* degradation of ruminant feeds. *Journal of Theoretical Biology*. 163: 99-111.) using the MLP (Ross,

G.J.S. 1987: *MLP*, *Maximum Likelihood Program Version 3.08*. Oxford Numerical Algorithms Group) package. The equation is in the form,  $Y = A\{1 - e^{[-b(t-1) - c(\sqrt{t}-\sqrt{t})]}\}$  where Y is the cumulative gas production (ml), A is the asymptote (i.e. gas pool), T is lag time, and b (h<sup>-1</sup>) and c (h<sup>-0.5</sup>) are decay rate constants. A combined fractional rate (h<sup>-1</sup>) of gas production ( $\mu$ ) was calculated as,  $\mu = b + c/2\sqrt{t}$ , where t is the incubation time (h).

It can be seen for Festuca arundiancea (denoted as BN in Figure 25) that cell cultures have a higher rate of digestion and cumulative gas production in the presence of FAE and that the addition of an exogenous xylanase further enhance the availability of fermentable carbohydrates. Similar results are found in FAE expressing cultures without added FAE. Fermentation rates are further increased compared with controls by the addition of exogenous FAE or xylanaase as these cultures expressing FAE have a reduced cell wall phenolic composition to controls Figures 26-28.

15

20

#### Example 9

FAE & xylanase transformed plants

Addition of exogenous xylanase (GC140) greatly increased FAE mediated release of phenolics from *Festuca* and *Lolium* leaves expressing *A. niger* FAE. See Figures 29-31 which show that phenolic release from leaf cell walls is increased in all FAE expressing plants on cell death and this is stimulated by xylanase irrespective of the targeting. Therefore expression of a fungal xylanase in plant cells is tested.

The FAE expression cassette is modified to comprise a fungal xylanase gene (either *T. reesei* or *A. niger*) to yield a FAE-xylanase expression cassette. The FAE-xylanase expression cassette is used to transform plant cells in a manner similar to those described in Example 3. The transformed cells are allowed to grow and are selected on an appropriate medium. The enzymes so expressed increase the availability of fermentable carbohydrates to a greater extent than the FAE expression cassette.

30

25

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will

be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

#### What is clalmed:

5

10

20

 A transgenic plant comprising an expression cassette comprising a promoter operably linked to a ferulic acid esterase encoding polynucleotide.

- 2. The plant of claim 1, wherein the polynucleotide is derived from Aspergillus niger.
- 3. The plant of claim 2, wherein the polynucleotide is FAE! from Aspergillus niger.
- The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with an altered glycosylation site.
  - 5. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with a substitution so that glycosylation is altered.
- 6. The plant of claim 3, wherein the polynucleotide further comprises a polynucleotide that encodes CTWPVAAA at the 3' end.
  - 7. The plant of claim 3 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.
  - 8. The plant of claim 1, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by sexual reproduction.
  - 9. The plant of claim 1, wherein the promoter is an inducible promoter.
  - 10. The plant of claim 9, wherein the promoter is a senescence promoter.
- 11. The plant of claim 9, wherein the promoter is a heat shock promoter.
  - 12. The plant of claim 1, wherein the promoter is a constitutive promoter
  - 13. The plant of claim 1, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the

polynucleotide.

10

15

20

25

14. The plant of claim 13, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

- 15. The plant of claim 14, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene
  - 16. The plant of claim 15, wherein the targeted gene is a barley aleurain gene.
  - 17. The plant of claim 15, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum or apoplast signal sequence.
  - 18. The plant of claim 15, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted senescence gene.
  - 19. The plant of claim 18, wherein the senescence gene is a Lolium See1 signal sequence.
- 20. The plant of claim 13, wherein the polynucleotide is derived from the signal sequence of a golgl targeted gene.
- 21. The plant of claim 20, wherein the targeted gene is a rat sialyl transferease signal sequence.
- 22. The plant of claim 13, wherein the polynucleotide is derived from the signal sequence of an apoplast signal sequence.
  - 23. The plant of claim 22, wherein the signal sequence is from Aspergillus niger ferulic acid esterase.
  - 24. The plant of claim 16, wherein the polynucleotide is derived from *Solanum tuberosum*.
  - 25. The plant of claim 13, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide
  - 26. The plant of claim 25, wherein the polynucleotide sequence is a KDEL sequence.

27. The plant of claim 25, wherein the polynucleotide sequence is a stop codon.

- 28. The plant of clalm 25, wherein the polynucleotide sequence is an extension of the ferulic acid esterase reading frame to provide a linker to KDEL.
- 29. The plant of claim 1, further comprising Introduction into the plant a second expression cassette comprising a promoter operably linked to a xylanase encoding polynucleotide.
- 30. The plant of claim 29, wherein the xylanase encoding polynucleotide is from *Trichoderma reesei*.
  - 31. The plant of claim 29, wherein the first and second expression cassettes are present on separate plasmids.
  - 32. The transgenic plant of claim 1, selected from the group consisting of Festuca, Lolium, Zea and Avena.
  - 33. The transgenic plant of claim 32, wherein the plant is a Festuca plant.

15

20

25 -

- 34. A method of controlling the level of phenolic acids in plant cell walls of a transgenic plant, the method comprising introducing into the plant an expression cassette comprising a promoter operably linked to a ferulic acid esterase encoding polynucleotide.
- 35. The method of claim 34, wherein the polynucleotide is derived from Aspergillus niger.
- 36. The method of claim 35, wherein the polynucleotide is a FAE 1 gene from *Aspergillus niger*.
- 37. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with an altered glycosylation site.
- 38. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with a substitution such that glycosylation is altered.
  - 39. The method of claim 36, wherein the polynucleotide

comprises CTWPVAAA at the 3' end.

40. The method of claim 36 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.

- 41. The method of claim 36, wherein the polynucleotide comprises SEQ ID NO:1.
  - 42. The method of claim 34, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by transformation of cell cultures.
- 43. The method of claim 42, wherein the cell cultures are regenerated to plants.
  - 44. The method of claim 34 wherein the ferulic acid esterase polynucleotide is introduced into the plant by sexual reproduction.
  - 45. The method of claim 34, wherein the transgenic plant is a member of a genus selected from the group consisting of Festuca, Lollum, Avena and Zea.
  - 46. The method of claim 45, wherein the transgenic plant is a member of the genus Festuca.
  - 47. The method of claim 46, wherein the transgenic plant is a Festuca arundinacea.
- 20 48. The method of claim 34, wherein the promoter is an inducible promoter.
  - 49. The method of claim 48, wherein the promoter is a senescence promoter.
  - 50. The method of claim 48, wherein the promoter is a heat shock protein promoter.
    - 51. The method of claim 34, wherein the promoter is a constitutive promoter.

25

52. The method of claim 51, wherein the promoter is an actin

promoter.

5

10

15

20

25

53. The method of claim 34, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the polynucleotide.

- 54. The method of claim 53, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.
- 55. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene.
- 56. The method of claim 55, wherein the targeted gene is a barley aleurain gene.
- 57. The method of claim 55, wherein the polynucleotide is derived from the signal sequence of a Lolium See1 signal sequence.
- 58. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum signal sequence.
  - 59. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce an apoplast signal sequence.
  - 60. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a golgi targeted gene.
  - 61. The method of claim 60, wherein the targeted gene is a rat sialyl transferease signal sequence.
- 62. The method of claim 59, wherein the polynucleotide is derived from the signal sequence of a fungal apoplast signal sequence.
- 63. The method of claim 62, wherein the signal sequence is from Aspergillus niger feruitc acid esterase.
  - 64. The method of claim 59, wherein the apoplast signal

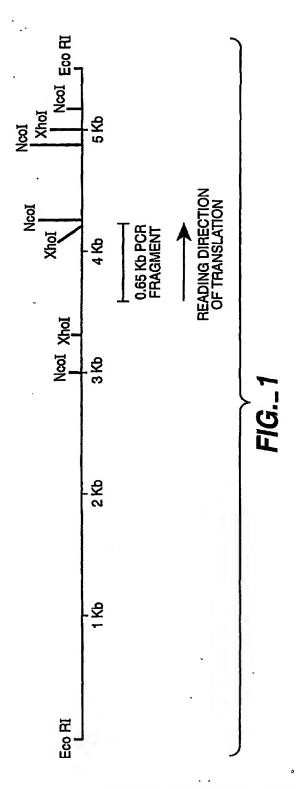
sequence is from a potato.

5

15

65. The method of claim 53, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

- 66. The method of claim 65, wherein the polynucleotide sequence is a KDEL sequence.
- 67. The method of claim 65, wherein the polynucleotide sequence is a stop codon.
- 68. The method of claim 65, wherein the polynucleotide
  sequence is an extension of the ferulic acid esterase reading frame to provide a linker to KDEL.
  - 69. The method of claim 34, further comprising simultaneous introduction into the plant a second expression cassette comprising a promoter operably linked to a polynucleotide encoding a xylanase gene.
  - 70. The method of claim 69, wherein the second polynucleotide is a fungal xylanase.
    - 71. The method of claim 70, wherein the fungal xylanase is from *Trichoderma reesei*.
- 72. The method of claim 35, wherein the first and second expression cassettes are present on separate plasmids.
  - 73. The method of claim 1, wherein the first and second expression cassettes are present on separate plasmids.
    - 74. A transgenic plant produced by the method of claim 34.



SUBSTITUTE SHEET (RULE 26)

Nco I	EcoR V	Psp1406		
CCATGGTGGTGTCGATAI	CGATATCGGCAGTAGTCT	TTGCCGAAACGTTGAGGGTTACAC	CGGCAGTAGTCTTTGCCGAAACGTTGAGGGTTACAGTGATCTGCGTCGGACATACTTCGGGGAATCTACGGC	ექ 80
GGAATATCAAAGTCTTC	TCTTCGGAATATCCATAT	TGGGAAAGGACAGAAGCTCCGGGG	Sac I SGAATATCCATATTGGGAAAAGGACAGAAGCTCCGGGGTAGTTTGATAGATGAGCTCCGGTGTATTAAATCGGG	. 081
GACAGGAG	TGAGCGTCATGTAGACCA	BSSH II	BSSH II AGCTGACAGGAGTGAGCGTCATCTAGTAATGTCAGTCGCGCGCAATTTCGCACATGAAACAAGTTGATTTCGGGACCCCAT	1 270
TGTTACATCTCTCGGCT/		BSt1107 I	Xho I BSt1107 I ECIHK I CAGCTCGAGATGTCCAAAAAGTCAGGG	09E 7
TGAAACGAT	Pvu I TCGTCGGATATTTCTTG	TITTATCCTAAATTAGTCTTCCAG	Pvu I AATATGAAACGATCGTCGGATATTTCTTGTTTTTTATTAGTCTTCCAGTGGTTTATTTA	T 450
CAACGGAC	TTCTCATACCACTCATT	BACATAATTTCAAACAGCTCCAGG	Xmn I CATCCAACGGACTTCTCATACCACTCATTGACATAATTTCAAACAGCTCCAGGCGCATTTAGTTCAACATGAAGCAATTCTCCGCCAAAC	C 240
		FIG2A	signal sequence	

1 810 720 Eco31 1 ATACTAACTACACCCTCACGCCTTTCGACACCCTACCACAATGCAACGGTTGTGAAGTACACGGTGGATATTATATTGGATGGGTCTCCG 900 630 TGGCCACTATCTCCCAAGCTGCCTACGCCGACCTGTGCAACATTCCGTCGACTATTATCAAGGGAGAAAATTTACAATTCTCAAACTG **ACATTAACGGATGGATCCTCCGCGACGACGACAGCAAAGAAATAATCACCGTCTTCCGTGGCACTGGTAGTGATACGAATCTACAACTCG** ACGTCCTCGCAGTTGTGGTGACTGCAGGCACGCCTTAGCAGCCTCTACGCAAGGCATCTCCGAAGACCTCTACAGCCGTTTAGTCGAAA w S BsaB I œ 0 တ w G ဟ ¥ œ G O Sall တ ⋖ Bpu101 w z ဟ ~ Pst ∢. signal sequence 4 œ ⋖ BamH 1 O ဟ G Z Msc ⋖

FIG.\_2E

	0	1080				2	<del>-</del> 0	1280			1960	000
BSPM I Acc III Acc III GCTTGTGCGCTGACGGCCACAGGTATGCCTTGTGACGGGCCACAGGTATGCCCTCG	Y A L T V T G H S	ApaB   Pvu    GTATAATACTCACTAACTCTAGGATCTCGGAGGGTCCCTGGCAGCACTCACT		t ats.	ccietacaccitceeceaaccececaceecaarcageccitcecetcetacateaaceatecctrccaae	Q A F A S Y M N D A F Q	BspM i Nco i	GCAGTATTTCCGGGTCACTCATGCCAACGACGGCATCCCAAACCTGCCCCCGGTGGAGCAGGGGTACGCCC	IPNLPPVEOGYA		GGAGCGTTGATCCTTACAGCGCCCAGAACACATTTGTCTGCACTGGGGATGAAGTGCAGTGCTGTGAGGCCC	CTGDEVOCCEA
Acc III ICAGCAGGTTAGCCAGTATCCGGACTA	0 0 4 5 0 7 P D	rcactaactctac6atagtctcgga6			OTTCGGCGAACCGCGCAGCGGCAATC	н В В В В В В В В В В В В В В В В В В В		CCGGGTCACTCATGCCAACGACGGCA	RVTHANDG		TCCTTACAGCGCCCAGAACACATTTG	PYSAONTFV
Tth1111 BspM I	0 V E S L V K	TGATTTCTTTCAATTAAGTGTATAATAC	Intron	BsrG	CGACATACGACAACATCCGCCTGTACAC	YDNIRLYT		CCTCGAGCCCAGATACGACGCAGTATTT	POTTO9	Sca I	<b>ATGGCGGTGTAGAGTACTGGAGCGTTGA</b>	VEYWSVD
TCCAGGAC	0 >	TGATTTCT			CGACATAC	<b>≻</b>	You'X	CCTCGAGC	S S		ATGGCGGT	9 9

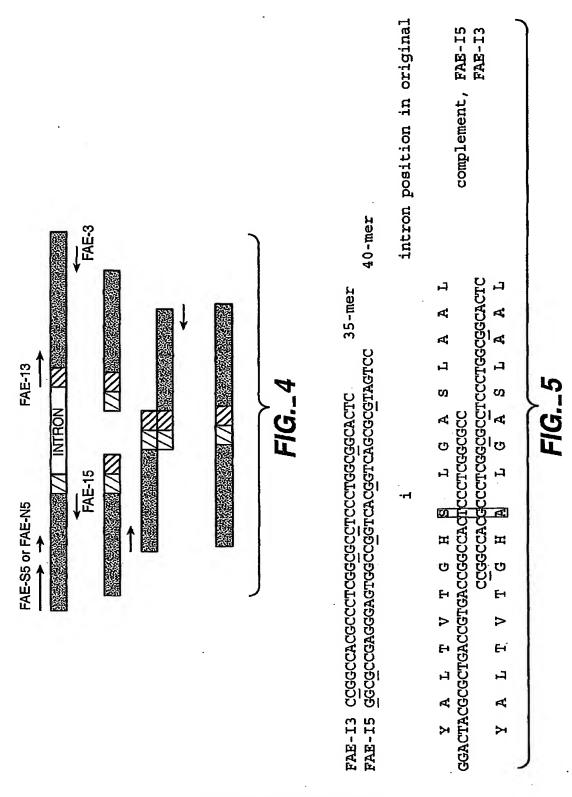
SUBSTITUTE SHEET (RULE 26)

1890	CTCAATTAT	TACCATGGT	GCTAGCTC	GAAGA	30CTGC	TGCATCGCTATATGATCCCATAAAGAAGCAACAACTTTCAGATCTCGTTTTGCGCTGCGAAGAGCTAGCT
	Eco31	- 00 N	Nhe -		Sap I	BfrB i Bgl II
1800	CAGCCTAAA	TATGAATAC	AGTACTAA	AAGGT	ATTCT	GTCCCGTTTAAATCAAACCTTTCAGCCTAGCACAGTCAGAATACACCCAATCTAAGGTAGTACTAAATATGAATACAGCCTAAA
Ppu10 !	<u> </u>		Scal			Drail
1710	AATGTATCA	ATGAGCGA/	CCTTGAATG	AGACC	GTGAT	TGACAGATATCTCTAAACACCTTATCCGCTTAAACCCATCATAGATTGTCACGTGATAGACCCCTTGAATGAGCGAAATGTATCA
						EcoR V
1620	i TAAAGTAGA	TATATATT	GATTGTGA/	CTTTT	TTTCT	AAGTTCCTTCCATGAATAGATATGGTTACCCTCACCATAAGCCTTGAGGTTGCCTTTCTCTTTTGATTGTGAATATATAT
	Dra (					BstEill
1530	ACATGTAAT	AAATCAGG/	TTTTTCGG	CACAC	CGAAG	CGAGTGTACCAGGAAAGATGGATGTCCTGGAGGGGCATGCAT
	BspLU111				1 201	Ppu10   BfrB   SnaB   Sph   Bst1107
2				ب ا	G A	G G O G V N A H T T Y F G M T S
1440	CAGCCTCCC	AGTCATTT	ATGGTGAT(	TGTAC	GAGCC	AGGGCGGACAGGGTGTGAATAATGCGCACACGACTTATTTGGGATGACGAGCGGAGCCTGTACATGGTGATCAGTCATTTCAGCCTCC
			Boll	BsrG	BsrB 1	Fsp   Bsi

FIG.\_2D

		BspLU11	Sma I
GAGTGGAGCGTTTAGTCTCGTTTA	AGCCTAGCTATCTTATAAGGACAAC	GTTTAAGCCTAGCTATCTTATAAGGACAACACATGTACATGGGCTTACTTGTAGAGAGGTAGGATCCCGGG	
Xho !	BseR1	Th111	008
CTTCTTCACATCTCGAGGAGTTGT	стасасвтсесвтскатвтсатаав	GTTGTCTACACGTCGCGTCCATGTCATAAGCCGGTACTCGACGTTGTCGTGACCGTGACCCAGACCCCTGT	CCAGACCCCTGT
		Nco I	
TGATAGCGTTGAGAGGCCCTATA	TTTGAATTTCCAATCTCAGCTTTAC	TGATAGCGTTGAGAGGCCCTATATTTGAATTTCCAATCTCAGCTTTACGAAGATATGCCCATGGTGGAGGGTTAGTAAACCGATGATGA	AACCGATGATGA 2160
Eco31	MscI	BspLU11	
TCGTGTGCAGCATGAGATGAGACC	GTGGCCAATCCTGTTCAAATGCCAA	AGACCGTGGCCAATCCTGTTCAAATGCCAAGACCCGCCTCCTACCACATGTAAGGCATCCGTCGGCCGCAC	CGTCGGCCGCAC 2250
		Xcm I Msc I	BsrD I
GTTGAATTGTGCAAATGCCGAGAT	CATAAAAGCGGCCACGT	GAGATCATAAAAGCGGCCACACTTCCACGTCGGTACTGGGTTGCGCGTGGCCATACTGTGTTTCCA 2340	TETETTTTCCA 2340
	Alwn (	Ear	Vsp I
TTGCGTGGGTCGTTCGTGTTACTG	CGACGCAGATICTGTAGGCAAGGCG	ACTGCGACGCAGATTCTGTAGGCAAGGCGCAGGGCTCTCTTCTGAGGTAGAAACACCCCATATTAATCT	CATATTAATCT 2430
Econ I			
GAATTC 2436	FIG.	FIG. 2E	
	· · ·		

**TCACGCCTTTCGACACCCTACCACAATGCAACGGTTGTGAAGTACACGGTGGATATTATATTGGATGGGTCTCCG** GCGGCAATCAGGCCTTCGCGTCGTACATGAACGATGCCTTCCAAGCCTCGAGCCCCAGATACGACGCAGTATTTCC **AGGGCGGACAGGGTGTGAATAATGCGCACACACTTATTTGGGATGACGAGCGGAGCCTGTACATGGTGATCAG** TOCATICGUTATATGATICCIATAAAGAAGCAACAAICTTTICAGATICTICGTTTTGCGCTGCGAAGAGCTAGCTIAI CATGGTCTCAATTATGAGTGGAGCGTTTAGTCTCGTTTAAGCCTAGCTATCTTATAAGGACAACACATATACATG CGGGGAATCTACGGCGGAATATCAAAGTCTTCGGAATATCCATATTGGGAAAGGACAGAAGCTCCGGGGTAGTTT GATAGATGAGCTCCGGTGTATTAAATCGGGAGCTGACAGGAGTGAGCGTCATGTAGACCATCTAGTAATGTCAGT CGCGCGCAATTTCGCACATGAAACAAGTTGATTTCGGGACCCCATTGTTACATCTCTCTGGCTACAGCTCGAGATG TOCCTGCCGAGTATACTTAGAAGCCATGCCAGCGTGTTGTTATACGACCAAAAGTCAGGGAATATGAAACGATCG **CATCCAACGGACTTCTCATACCACTCATTGACATAATTTCAAACAGCTCCAGGCGCATTTAGTTCAACAGG AATTCTCCGCCAAACACGTCCTCGCAGTTGTGGTGACTGCAGGGCACGCCTTAGGAGCCTCTACGCAAGGCATCT** CCGAAGACCTCTACAGCCGTTTAGTCGAAATGGCCACTATCTCCCAAGCTGCCTACGCCGACCTGTGCAACATTC CGTCGACTATTATCAAGGGAGAGAAATTTACAATTCTCAAACTGACATTAAACGGATCCTCCGCGACGACA **TCCAGGACCAAGTCGAGTCGCTTGTCAAACAGGATTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC** acaggtatgccctcgtgatttctttcaattaagtgtataattaactcactaactctacgatggtccc TGGCAGCACTCACTGCCGCCCAGCTGTCTGCGACATACGACATCCGCCTGTACACCTTCGGCGAACCGCGCA **GOGTCACTCATGCCAACGACGGCATCCCAAACCTGCCCCCGGTGGAGCAGGGGTACGCCCATGGCGGTGTAGAGT** ga ggttgcctttctcttttgattgtgaatatatattttaaagtagacagatatctctaaaacaccttatccgct aaccctttcagcctagcacagaatacaccaacccattctaaggtagtactaantatgaatacagcctaaa **GCCTTACTTGTAGAGGTAGGATCCCGGGCTTCTTCACATCTCGAGGAGTTGTCTACACGTCGCGTCCATGTCA TAAGCCGGTACTCGACGTTGTCGTGACCGTGACCCCAGACCCTGTTGATAGCGTTGAGAGGCCCCTATATTTGAA** TTTCCAATCTCAGCTTTACGAAGATATGCCCATGGTGGAGGGTTAGTAAACCGATGATGATCGTGTGCAGCATGA GATGAGACCGTGGCCAATCCTGTTCAAATGCCAAGACCCGCCTCCTACCACATGTAAGGCATCCGTCGGCCGCAC GITGAATTGTGCAAATGCCGAGATCATAAAAGCGGCCACACTTCCACGTCGGTACTGGATGGGTTGCGCGTGGCC **ATACTGTGTTTTCCATTGCGTGGGTCGTTCGTGTTACTGCGACGCAGATTCTGTAGGCAAGGCGCAGGGCTCTCT** CTGAGGTAGAAACACCCCATATTAATCTGAATTC



**SUBSTITUTE SHEET (RULE 26)** 

linker + stop codon linker + frameshift linker + KDEL C terminal none C-signal + codon optimisation Amp ± active glyc site ± 32aa clip site Gene ± ser to ala Vector Construction FAE gene nos sialyl tranf golgi signal aleurain apo signal aleurain vac signal FAE excretion N-signal pt none N terminal Promoter CaMV35S actin + intron senescence heat shock

SUBSTITUTE SHEET (RULE 26)

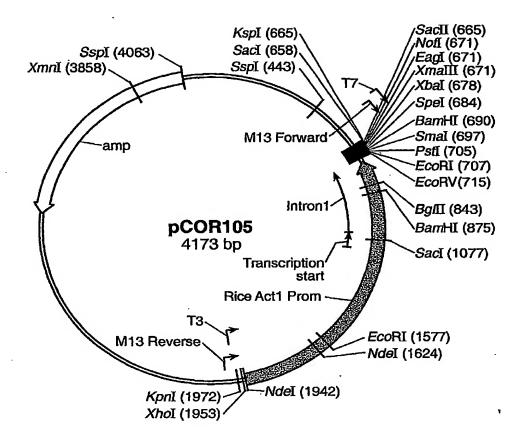
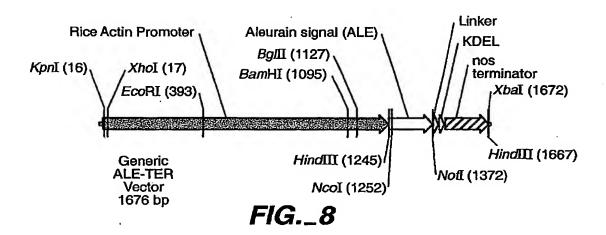


FIG.\_7



KDEL-COOH ER retention sequence

NotI

A A A K P L K D E L \*

1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

FIG.\_9

#### FAE-LINKER-FRAMESHIFT Structure and Sequence

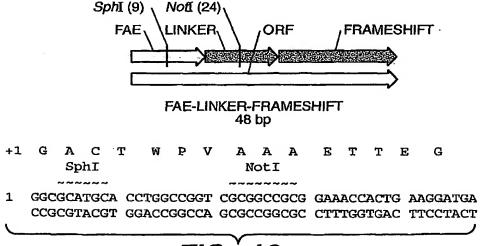


FIG.\_10

See1	1	UB8.1	•	•	•		•	•	,	,	•	,	PJQ5.2			FIG 11	
Ę.	1	CK3	UK12	UK13	UK6	UC5.1	UK2	UH10	UH11	UK1			•	•			]
Actin	(+hyg)	OH4	OH6	UH7	SHO HH	НОХЗ	UH3	UH8	OH9	UF1	•	pJQ4.9*	pJQ3.2*	pJO6.3*		Linker t KPLKDEL	
	Target	VAC	APO	APO	VAC	VAC	VAC/E.R.	E.R.	œ	E.R.	APO	GOLGI	APO	VAC	CO Site)	Linker Frameshiff	
σ			1					1	Í				i	1	stored N	Stop codon	
Plant Transformation Cassettes		WILLIAM STATES	**************************************	WITTING	MILLIAM		VIIIIII 8888	一							* - Modified Actin Promoter (Kpn1-EcoR1 Deletion and Restored NCO Site)	Aspergillus FAE Signal	
Plan	- hyg	T	1	Y	Y	Y		Y	Ţ				1	Ĭ	omoter (K	PPI Asp	
	Orlginal Actin + hyg	TR9.4	TR5.5	•	•	TR8 (glycos)			ı	TR6.1	TR2				lodified Actin Pr	RST Signal	
/ectors	£	<b>E</b>		•		•			1	TT2					*	Aleurain NPGR	
Initial Vectors	Original Actin	TP11.1	TT5	UA4.4	TP8.5	TP3.1	T04	TUS	ഉ	TP5.1	TP4	TP3.1	TP3.1	TP3.1		Aleurain NPIR	

SUBSTITUTE SHEET (RULE 26)

#### Vectors

### Original Actin promoter in pCOR105

·			
	Target	Signal sequences	Vectors
(i)	APO	- alcurain-NPGR-FAE - alcurain-dcINPIR -FAE	pUH6, pTT5, TT5.5, pTT5.1 pUH7, pUA4.4,
(ii) ·	ER	- aleurain-NPGR-FAE-linker-KDEL - aleurain-delNPIR-FAB-linker-KDEL	pTUS, pUH8, pUG4, pUH9,
(iii)	VAC	- aleurain-NPIR-FAE	pTP11.1, pTR9.4, pUH4, pUK3,
(iv) (v)	ER/VAC VAC	- aleurain-NPIR-FAE-linker-KDEL - aleurain-NPIR-FAE-linker-frameshift	pTU4, pUH3, pUA1K3, pTP3.1, pUC5.11
(vi) (vii)	VAC ER	- aleurain-NPIR-FAE-linker-stop - Aspergillus signal -FAE-KDEL	pTP8.5, pUH5 pTP5.1, pTP6.1, pUF1,
	ied actin promo	oter (Kpn1-EcoR1 deletion and restored	
(i)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pJ06.3
(ii) (iii)	GOLGI APO	- RST-FAE-linker-frameshift - PPI-FAE-linker-frameshift	pJQ3.2
(ui)	AO	- FFI-FAD-linker-framesimt	pJQ4.9
Heat-	shock promot	er	
<b>(i)</b>	APO	- aleurain-NPGR-FAE - aleurain-deINPIR-FAE	pUH12 pUH13

<b>(i)</b>	APO	- aleurain-NPGR-FAE - aleurain-deINPIR-FAE	pUH12 pUH13
(ii)	ER	<ul> <li>Aspergillus signal-FAE</li> <li>aleurain-NPGR-FAE-linker-KDEL</li> <li>aleurain-delNPIR-FAE-linker-KDEL</li> </ul>	pTP4a2, pTR2.22, pUH10 pUH11
(iii)	VAC	- aleurain-NPIR-FAE - aleurain-NPIR-FAE-linker-KDEL	pUK3,pTT3
(iv)	ER/VAC		pUK2
(v)	VAC	<ul> <li>aleurain-NPIR-FAE-linker-frameshift</li> <li>aleurain-NPIR-FAE-linker-stop</li> <li>Aspergillus signal -FAE-KDEL</li> </ul>	pUC5.11, pHOX3
(vi)	VAC		pUK6
(vii)	ER		pUK1, pTT2

### Senescence promoter

(i)	APO	- See1-PPI-FAE-linker-frameshift	pJQ5.2
(ii)	VAC	- See1-aleurain-deleted NPIR-FAE	pUB8.1

FIG.\_12

~~~~~

### ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast) Structure and Sequence

NPIR Underline
NPGR Bold

NcoI (9)
HindIII (2)
ORF
NPGR

ALEURAIN-NPIR 134bp

- +1 MAHARVLLLALAVLATAAVA
  HindIII Ncoi
- 1 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGGCACGA CCGGTGCCGG CGGCAGCGGC
- +1 V A S S S F A D S N P I R P V T D R A A . Noti

FIG.\_13

RAT SIALYL TRANSFERASE Golgi signal sequence

Hindill

AAGCTTACCA TOATCCACAC CAACCTCAAA AAGAAGTTCT CCTCTTCAT CCTCGTCTTC CTCCTCTTCG r r L L Ø Н

· V I C V W K K G S D Y E A L T L Q A K E F Q M CCGTGATCTG CGTGTGGAAG AAGGGCTCCG ACTACGAGGC CCTCACCTC CAAGCCAAGG AGTTCCAAT 71

26236255 \*\*\*\*\*\*\*

Noti

POTATO PROTEASE INHIBITOR II Apoplast signal sequence

Hindill

AAGCTTACMA TGGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTACCTCCT GATCGTSCTC Ø ц ø, z B × х о

GGCCTCTCT

12111, Ncol

M × H U ⋖. M 4 A 四 14 × **1** 

H

Q U TGCICGISIC CGCCAIGGAG CACGIGGACG CCAAGGCCIG CACCCKCGAG IGCGGCAACC TCGGCTTCGG 71

Ü

Noti

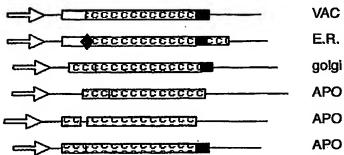
д С

CATCTGCCCG GCGGCCGCC

141

#### Targeting Expression of gfp to Different Cell Compartments

**Actin Promoter Targeting Vectors** 



Actin Promoter

Aspergillus signal

Aleurain signal (NPIR)

Aleurain signal I (del – NPIR)

📺 Aleurain signal (NPGR) 🖼

linker + stop codon

linker + KPLKDEL

Potato protease inhibitor

silyl transferase

FIG.\_16A

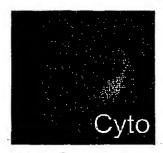


FIG.\_16B

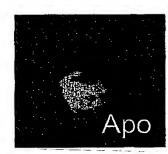


FIG.\_16C

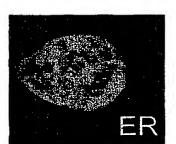


FIG.\_16D

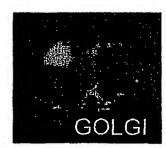


FIG.\_16E

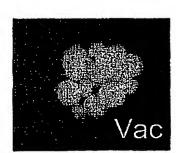


FIG.\_16F

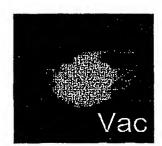


FIG.\_16G

# FAE Activity in Transgenic Festuca arundinacea Leaves of Different Ages Under ER and APO Targeting Sequence

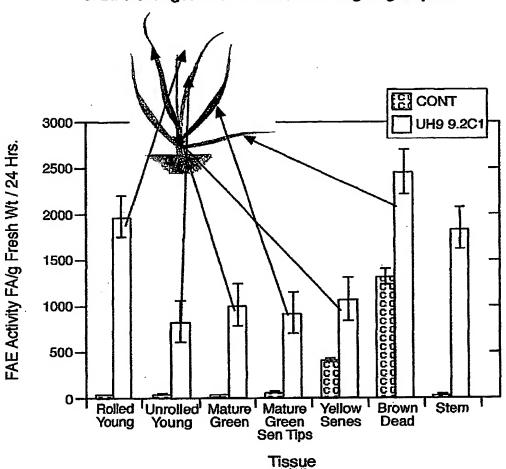


FIG.\_17A

# FAE Activity in Transgenic Festuca arundinacea Leaves of Different Ages Under ER and APO Targeting Sequence

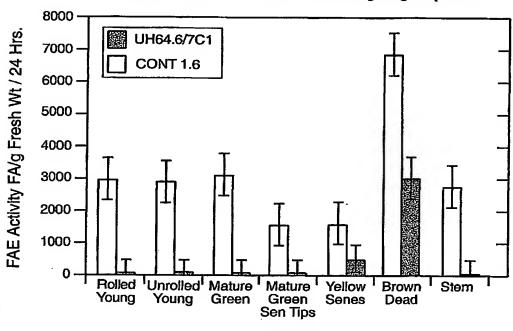
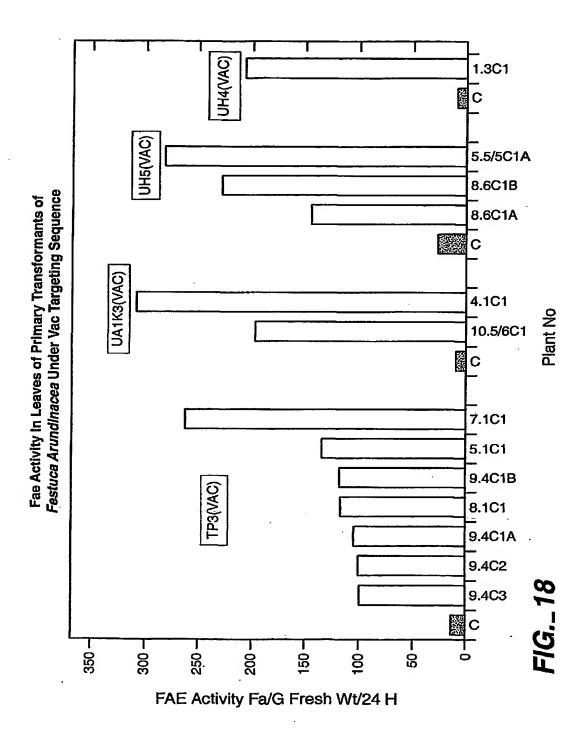
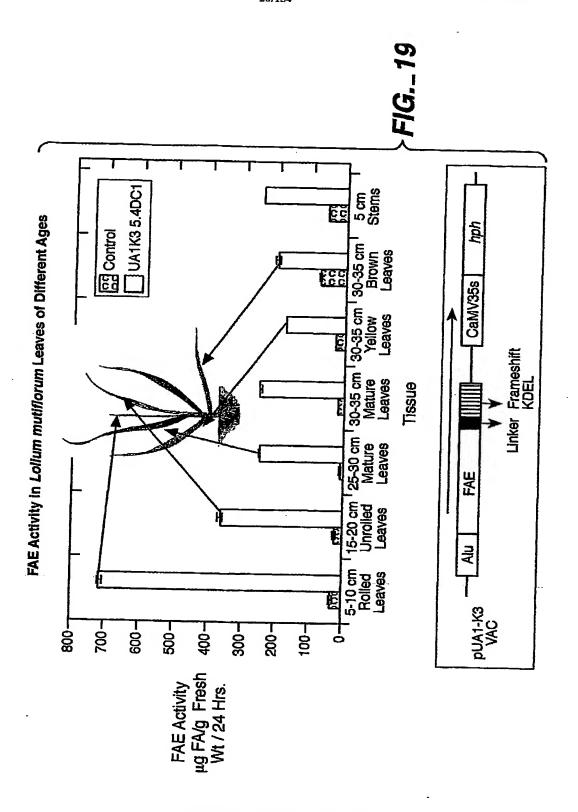


FIG.\_17B

Tissue

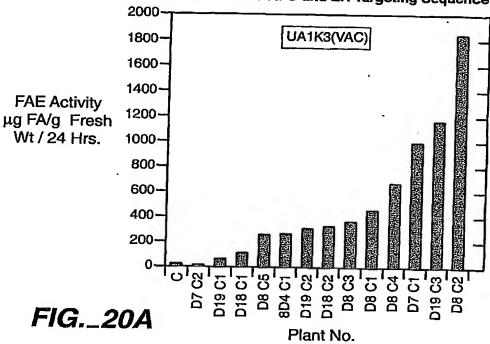


SUBSTITUTE SHEET (RULE 26)

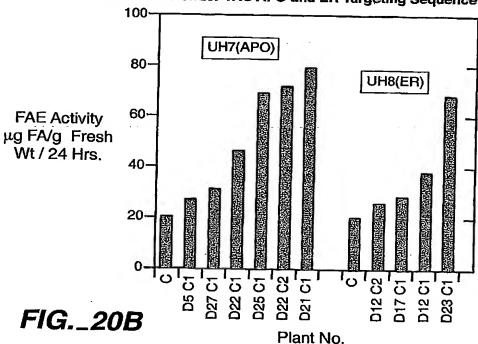


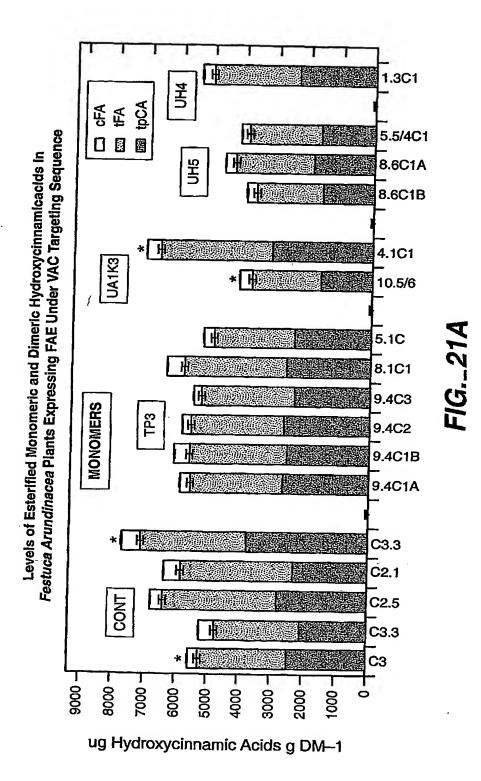
SUBSTITUTE SHEET (RULE 26)

FAE Activity in Leaves of Primary Transformants of *Lollum* multiflorum Under VAC APO and ER Targeting Sequence

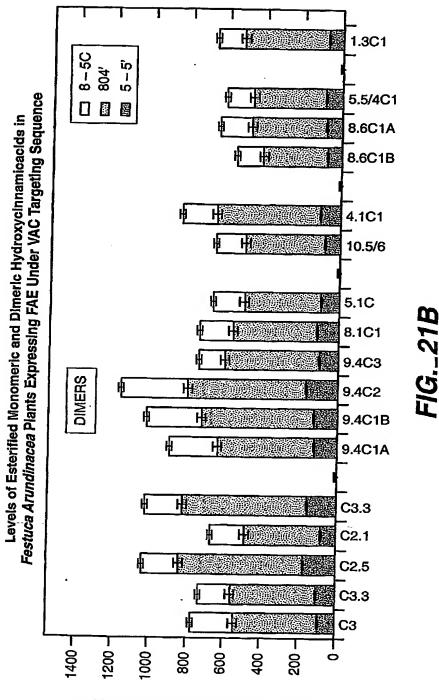


FAE Activity in Leaves of Primary Transformants of *Lolium* multiflorum Under VAC APO and ER Targeting Sequence

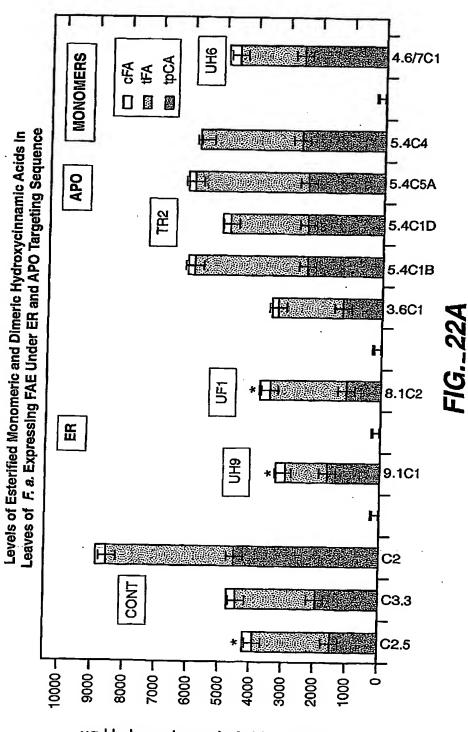




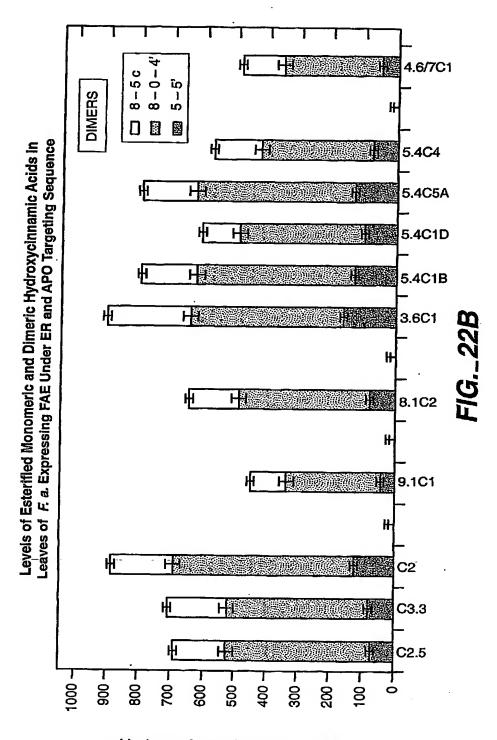
SUBSTITUTE SHEET (RULE 26)



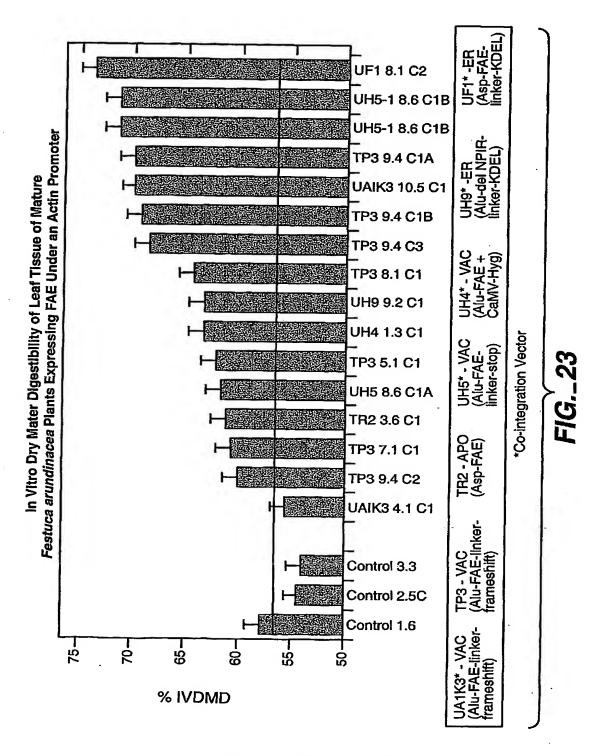
ug Hydroxycinnamic Acids g DM-1



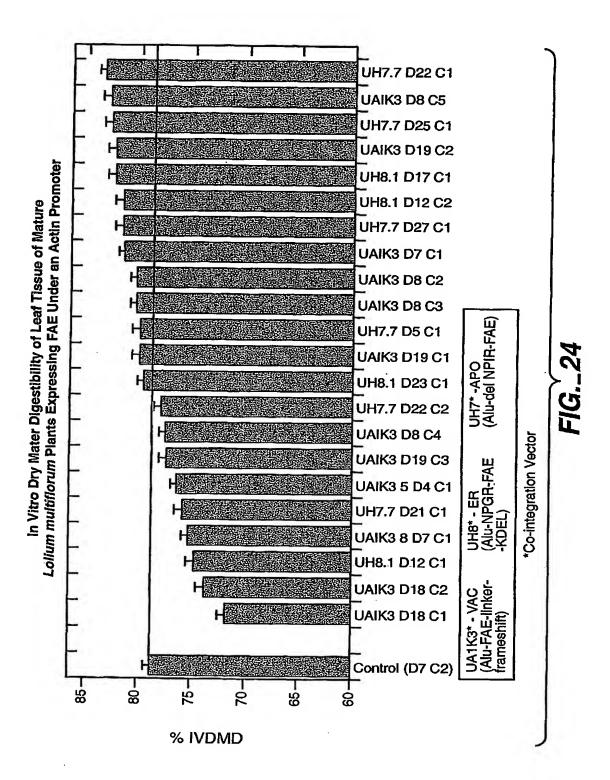
ug Hydroxycinnamic Acids g DM-1



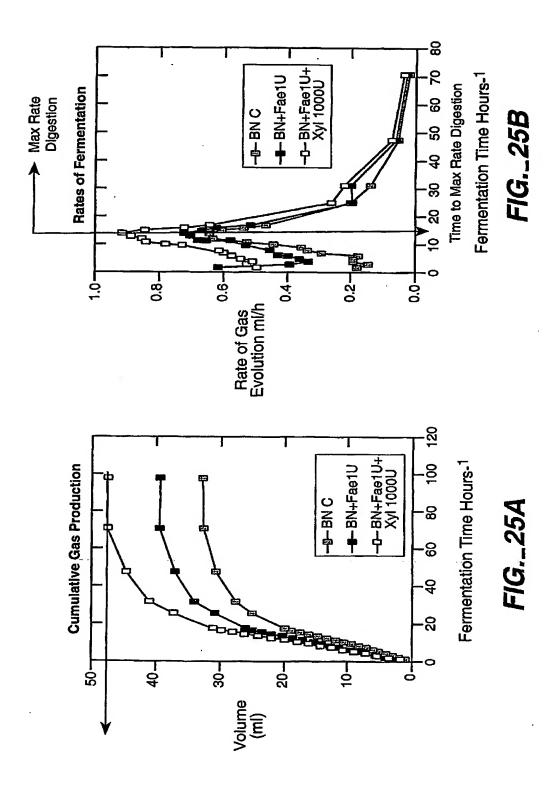
ug Hydroxycinnamic Acids g DM-1



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

In-vitro Fermentation of Festuca arundinacea Cell Walls From Cell Cultures Expressing Recombinant FAE1

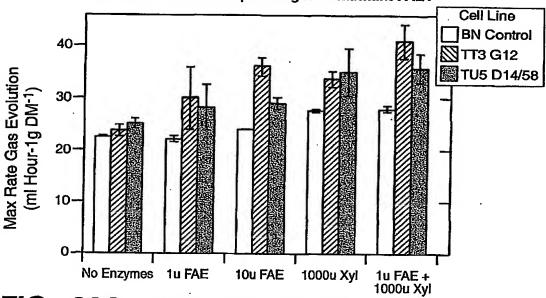
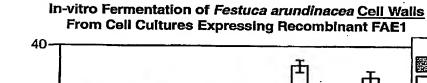


FIG.\_26A Maximum Rate of Digestion



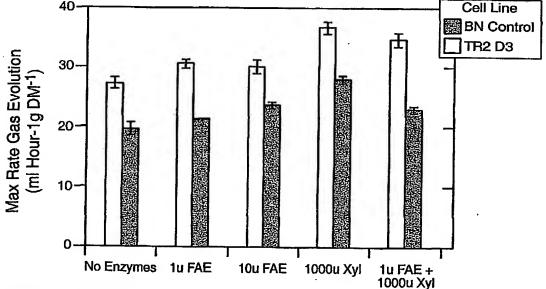


FIG.\_26B Maximum Rate of Digestion

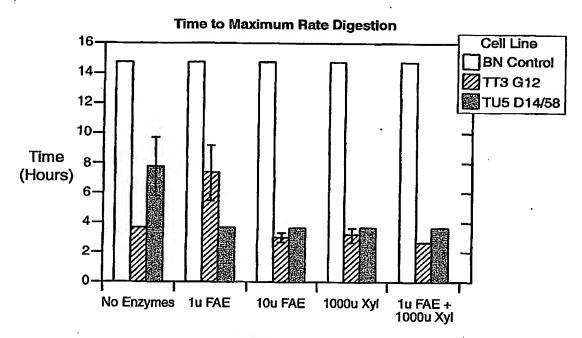


FIG.\_27A

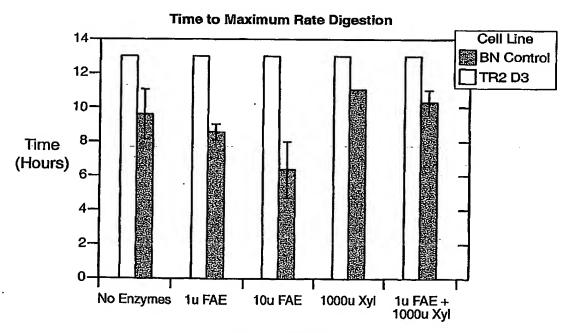
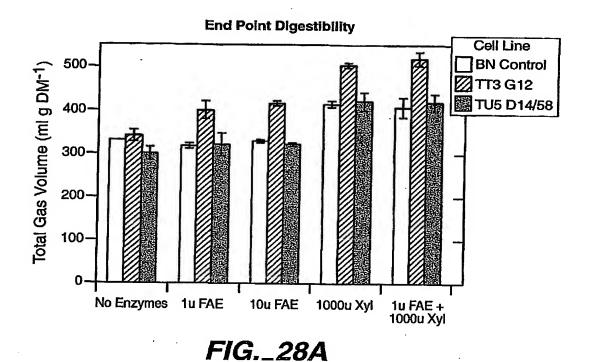


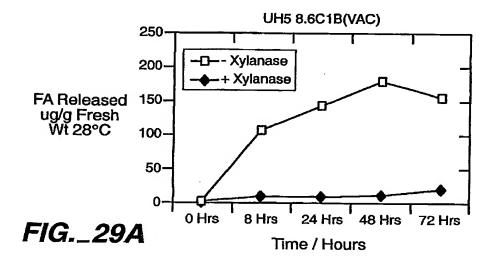
FIG.\_27B



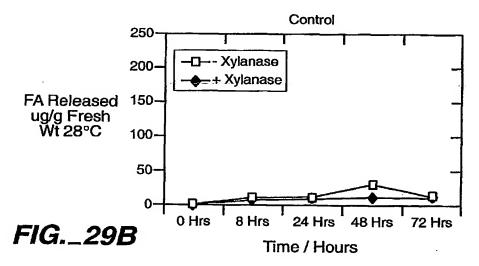
**End Point Digestibility** Cell Line 400-BN Control Total Gas Volume (ml g DM-1) 350 TR2 D3 300 250 200 150-100 50 No Enzymes 1u FAE 1000u Xyl 1u FAE + 1000u Xyl 10u FAE

FIG.\_28B

#### Kinetics of FAE Activity by Ferulic Acid Release from Cell Wall under Self Digestion in Festuca arundinacea and Stimulation by Xylanase



#### Kinetics of FAE Activity by Ferulic Acid Release from Cell Wall under Self Digestion in *Festuca arundinacea* and Stimulation by Xylanase.



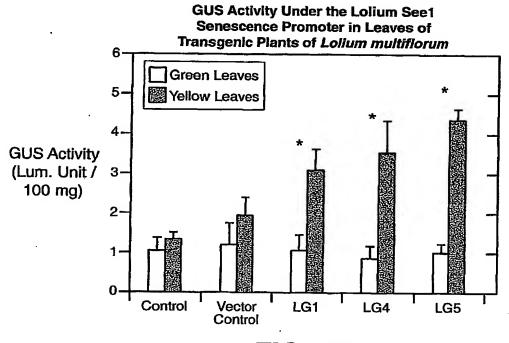
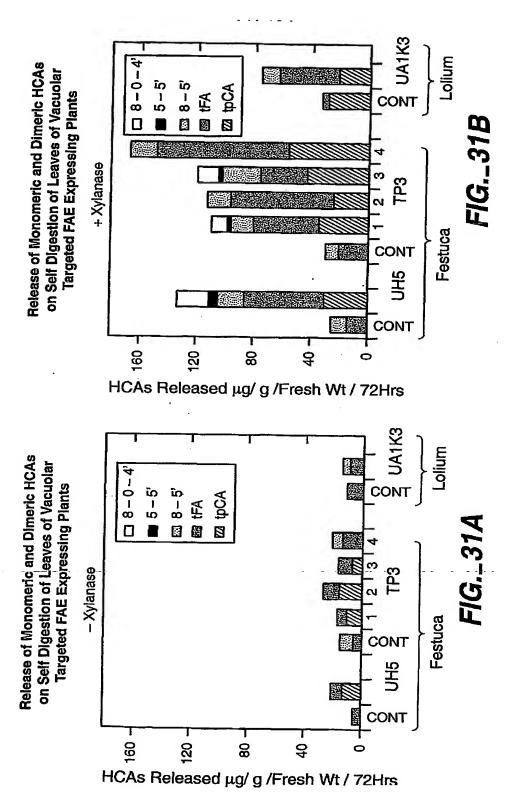


FIG.\_30



SUBSTITUTE SHEET (RULE 26)

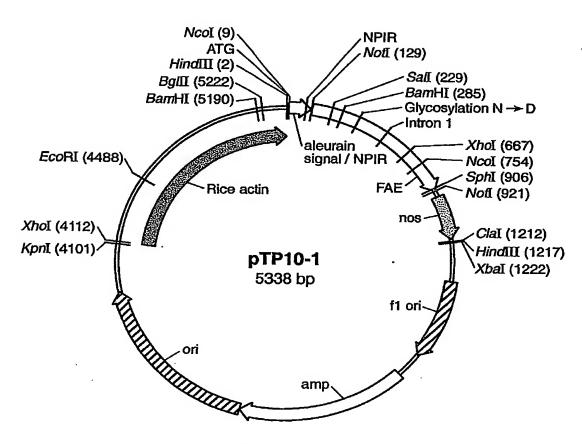


FIG.\_32A

Hindili 422222

-1

AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG 4 Н 4 ц 4 Н Н > 4 Ħ ď

NPIR

Noti

· A S S TCGCCTCCTC H

GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC

o:

141

SalI

D L C N I P S T I I K G E K I Y N B Q T D I N (GACCTGTGCA ACATTCCGTC GACTATATC AAGGGAGAGA AAATTTACAA TTCTCAAACT GACATTAACG

BamHI

\*\*\*\*\*\*

GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA GTGATACGAA Ω Ø O H U œ Ŀ Н ĸ ĽQ. Ø A A ĸ Н н 281

Glycosylation

TCTACAACTC GATACTGACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA CACGGIGGAT ATTATATIGG ATGGGICTCC GICCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGITA GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG 回 O ט M Q! ₹ D E E O H D E E E 畔 E 4 E 1 I I I A A A H ≯ O O ы • 351 421 491 561

211

GACGCAGTAT TTCCGGGTCA Qi GCCCAGATAC Ø Xhor Ø CGATGCCTTC Ě٤ 4 Α CGTACATGAA Ø GCCTTCGCGT 631

NdoI

2222222

GCCCATGGCG GTGTAGAGTA Þ Ö Ħ А GCAGGGGTAC O CCCCGGTGGA CCAAACCTGC Z CGACGGCATC Q

CTCATGCCAA

701

Ħ

GTGCTGTGAG TGCACTGGGG ATGAAGTGCA GATCCTTACA GCGCCCAGAA CACATTTGTC CTGGAGCGTT

Sphi

Ö GACGAGCGGC ಭ GAATAATGCG CACACGACTT ATTTTGGGAT Ħ 0 跆 H H 耳 4 Z Z GCCCAGGGCG GACAGGGTGT > Ø O Ø Ø ø

Noti

四四

TTAGAGICCC CATTTGGCAA TTGAATTACG TGATTATCAT ATAATTTCTG AAAGAAGCAG ATCGTTCAAA GTTTTTATGA CGCGCAAACT TATGAGATGG ACCACTGAAG GATGAGCTGT CGATAGAAA CAAAATATAG TGACGTTATT GGTCTTGCGA TCCTGTTGCC ATGTAATGCA GGCCGCGGAA TAAGATTGAA AATAATTAAC GCAATTATAC ATTTAATACG TTAAGCATGT GGCCGGTCGC TAAAGTTTCT 1051

AGGATAAATT

Hindill

\*\*\*\*\*\* 222222

ACCGATCGCC CCTATAGTGA CGTTACCCAA TCCAATTCGC AGAGGCCCGC AAAACCCTGG GTAATAGCGA CCGGTGGAGC CGTGACTGGG GCCAGCTGGC TCTAGAGCGG TTTACAACGT TCCCCCTTTC TCGATAAGCT TOGCCGTCGT TTGCAGCACA TGTTACTAGA GCGCGCTCAC GTGTCATCTA GTCGTATTAC CTTAATCGCC 1331 1261

CTGAATGGCG AATGGGACGC GCCCTGTAGC

GTTGCGCAGC

CTTCCCAACA

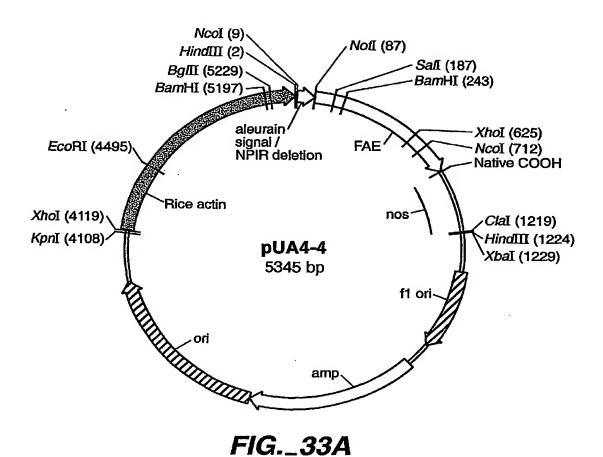
1401

GGCGCATTAA

| CGCTTTCTTC | TTAGGGTTCC | GTGGGCCATC | CTTGTTCCAA | ATTTCGGCCT | CGCTTACAAT | CATTCAAATA | GTATGAGTAT | TCACCCAGAA | CTGGATCTCA | TTAAAGTTCT | ACACTATTCT | GTAAGAGAAT | TCGGAGGACC | GGAACCGGAG | ACGLTGCGCA | AGGCGGATAA | TGGAGCCGGT | GTAGTTATCT | CCTCACTGAT | TCATTTTAA  | GAGTTTTCGT | TGCGCGTAAT | GCTACCAACT | TAGCCGTAGT | TACCAGTGGC | TAAGGCGCAG | GAACTGAGAT | ATCCGGTAAG | TTATAGTCCT | AGCCTATGGA         | TGTTCTTTCC | TCGCCGCAGC | CCGCCTCTCC | GCAGTGAGCG |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------------|------------|------------|------------|------------|
| CCGCTCCTTT | GGGGCTCCCT | GGTTCACGTA | ATAGTGGACT | GATTTTGCCG | AAATATTAA  | TTTCTAAATA | AAAAGGAAGA | CTGTTTTTGC | TTACATCGAA | ATGAGCACTT | GTCGCCGCAT | TGGCATGACA | CTGACAACGA | TTGATCGTTG | AATGGCAACA | GACTGGATGG | CTGATAAATC | CTCCCGTATC | GACATAGGTG | ATTTAAAACT | CCCTTAACGT | CCTTTTTTC  | CGGATCAAGA | CCTTCTAGTG | CTAATCCTGT | AGTTACCGGA | GACCTACACC | GCGGACAGGT | CCTGGTATCT | <b>AGGGGGGGCGG</b> | TTTGCTCACA | CTGATACCGC | AATACGCAAA | TGGAAAGCGG |
| GCCCTAGCGC | CTCTAAATCG | TTAGGGTGAT | ACGTTCTTTA | ATTTATAGG  | GAATTTTAAC | TTTGTTTATT | ATAATATTGA | TTTTGCCTTC | CACGAGTGGG | TTTTCCAATG | GAGCAACTCG | ATCTTACGGA | CAACTTACTT | GTAACTCGCC | TGCCTGTAGC | ACAATTAATA | TGGTTTATTG | ATGGTAAGCC | ACAGATCGCT | CTTTAGATTG | TGACCAAAAT | TTCTTGAGAT | GITTGITTGC | CAAATACTGT | CCTCGCTCTG | TCAAGACGAT | TGGAGCGAAC | AGGGAGAAAG | GGGGGAAACG | GATGCTCGTC         | TIGCIGGCCI | TTTGAGTGAG | AAGAGCGCCC | GTTTCCCGAC |
| ACTTGCCAGC | CCCCGTCAAG | AAAAACTTGA | GTTGGAGTCC | TATTCTTTG  | AATTTAACGC | GGAACCCCTA | AAATGCTTCA | TTTTGCGGCA | CAGTTGGGTG | CCGAAGAACG | CGCCGGGCAA | ACAGAAAAGC | ACACTGCGGC | GGGGGATCAT | GACACCACGA | CTTCCCGGCA | TCCGGCTGGC | CTGGGGCCAG | AACGAAATAG | CTCATATA   | GATAATCTCA | TCAAAGGATC | ACCAGCGGTG | GCGCAGATAC | CGCCTACATA | CGGGTTGGAC | CAGCCCAGCT | CGCTTCCCGA | GGAGCTTCCA | CGATTTTTGT         | TCCTGGCCTT | TATTACCOCC | GAGGAAGCGG | GGCACGACAG |
| TGACCGCTAC | CGCCGGCTTT | CTCGACCCCA | GCCCTTTGAC | TATCTCGGTC | ATTTAACAAA | AAATGTGCGC | TAACCCTGAT | TTATTCCCTT | TGCTGAAGAT | AGTTTTCGCC | CCCGTATTGA | CTCACCAGTC | ATGAGTGATA | TGCACAACAT | CGACGAGCGT | CTTACTCTAG | GCTCGGCCCT | CATTGCAGCA | ACTATGGATG | ACCAAGITTA | GATCCTTTTT | GTAGAAAGA  | AACCACCGCT | CTTCAGCAGA | TCTGTAGCAC | CGTGTCTTAC | TTCGTGCACA | GAMAGCGCCA | AGCGCACGAG | ACTTGAGCGT         | TTTTACGGT  | TGGATAACCG | GTCAGTGAGC | TAATGCAGCT |
| ACGCGCAGCG | TCGCCACGTT | TTTACGGCAC | ACGGTTTTTC | CACTCAACCC | AAATGAGCTG | CTTTTCGGGG | CATGAGACAA | CGTGTCGCCC | AAGTAAAGA  | GATCCTTGAG | GCGGTATTAT | TGGTTGAGTA | TGCCATAACC | ACCGCTTTTT | CCATACCAAA | TGGCGAACTA | CCACTTCTGC | CTCGCGGTAT | GAGTCAGGCA | TAACTGTCAG | TCTAGGTGAA | GTCAGACCCC | CAAACAAAAA | AGGTAACTGG | CTTCAAGAAC | GGCGATAAGT | GAACGGGGGG | TGAGCTATGA | GGAACAGGAG | GCCACCICTG         | CAACGCGGCC | CCTGATTCTG | AGCGCAGCGA | GCCGATTCAT |
| TGTGGTGGTT | CCTTCCTTTC | GATTTAGTGC | GCCCTGATAG | ACTGGAACAA | ATTGGTTAAA | TTAGGTGGCA | TGTATCCGCT | TCAACATITC | ACGCTGGTGA | ACAGCGGTAA | GCTATGTGGC | CAGAATGACT | TATGCAGTGC | GAAGGAGCTA | CTGAATGAAG | AACTATTAAC | AGTTGCAGGA | GAGCGTGGGT | ACACGACGGG | TAAGCATTGG | TTTAAAAGGA | TCCACTGAGC | CTGCTGCTTG | CTTTTTCCGA | TAGGCCACCA | TGCTGCCAGT | CGGTCGGGCT | ACCTACAGCG | CGGCAGGGTC | GTCGGGTTTC         | AAAACGCCAG | TGCGTTATCC | CGAACGACCG | CCGCGCGTTG |
| 1471       | 1541       | 1611       | 1681       | 1751       | 1821       | 1891       | 1961       | 2031       | 2101       | 2171       | 2241       | 2311       | 2381       | 2451       | 2521       | 2591       | 2661       | 2731       | 2801       | 2871       | 2941       | 3011       | 3081       | 3151       | 3221       | 3291       | 3361       | 3431       | 3501       | 3571               | 3641       | 3711       | 3781       | 3851       |

| TTTTGTGGTA<br>GCAGCCTCGT | TCTTTCTTCT TTTTGTGGTA<br>TGTGACAAAT GCAGCCTCGT | GGAATGGGGC TCTCGGATGT AGATCTTCTT TGTTCATCGG TAGTTTTTCT TTTCATGATT | TCTCGGATGT<br>TAGITITITCT | GTCGGCCCGG ATCCTCGCGG GGAATGGGGC TCTCGGATGT AGATCTTCTT GAATTTGAAT CCCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTCATGATT GCGGAGCTTT TTTCATGATT | ATCCTCGCGG<br>CCCTCAGCAT<br>TTTGTAGC | GTCGGCCCGG ATCCTCGC<br>GAATTTGAAT CCCTCAGC<br>GCGGAGCTTT TTTGTAGC                 | 5181<br>5251<br>5321 |
|--------------------------|------------------------------------------------|-------------------------------------------------------------------|---------------------------|-------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|-----------------------------------------------------------------------------------|----------------------|
|                          |                                                | Bglii                                                             |                           |                                                                                                                                     | Bambi                                | Ba i                                                                              |                      |
| CCGGGCGTGA               | GCTGGCGTCT                                     | GGATCTCGCG                                                        | GGGAGGGGCG                | ATCGGTGCGC                                                                                                                          | AGAGCGGCTT CGTCGCCCAG                | AGAGCGGCTT                                                                        | 5111                 |
| TGGGTGGGCG               | CTTGGTAGTT                                     | GATCTTTGGC                                                        | TCTCGGTCTC                | TTTTTTCG                                                                                                                            | TTTCTCCGTT                           | CICCICIFIC                                                                        | 5041                 |
| ECCCGCCCCT               | CCGGTAACCA                                     | TCCGCCGCCG                                                        | CCCCICCCCC                | GAGCTCCTCC                                                                                                                          |                                      | CCCCCTCGCT                                                                        | 4971                 |
| CCACCTCCTC               | ACCACCACCA                                     | CCCTACCACC                                                        | ATCCCCCCAA                | ATACCCCCC CTCTCCTCCC                                                                                                                |                                      | CACTATATAC                                                                        | 4901                 |
| CCCCCATCGC               | AAAGAAACGC                                     | CICCGCIFICC                                                       | CCGGCCCICC                | ATCGCGAGCA GCGACGAGGC                                                                                                               | ATCCCCGAGCA                          | AGCGAGGAGG                                                                        | 4831                 |
| GGGCCGGAAA               | CGGGTCGTGG                                     | CAGGTGGGTC                                                        | AGAAAAACAG                | AAAAAAAA                                                                                                                            | AAGAAAGAAA                           | AAAAAAAAA                                                                         | 4761                 |
| CTCGCAGCCA               |                                                | TTGTCCGCAC                                                        | TCACCGTGAG                | CCCGGCACTA                                                                                                                          | AGTCTCCACC                           | ACTGGCAAAT                                                                        | 4691                 |
| AGTGCAGCCA               | AACCCACCC                                      | AACCCAACCC                                                        | AGCCCAACCC                | TAGCAAGCCC                                                                                                                          | GCACGATCCA                           | AAAGTGCTAT                                                                        | 4621                 |
| CCTAAAGCCC               | CCCCTAAAGT                                     | ACAAAAACAA                                                        | CAAAACATTT                | ATAAAAGATA AACTTAGACT CAAAACATTT                                                                                                    | ATAAAAGATA                           | TTCTAGTAAA                                                                        | 4551                 |
| TGGGTATTT                | GTTGCAGCGA                                     | GCTTGCCCCC                                                        | GATTAAAATA                | TCAGGCGAAT TCCACAATGA ACAATAATAA GATTAAAATA                                                                                         | TCCACAATGA                           | TCAGGCGAAT                                                                        | 4481                 |
|                          |                                                |                                                                   |                           |                                                                                                                                     |                                      | 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1                                           |                      |
|                          |                                                |                                                                   |                           |                                                                                                                                     |                                      | ECOR                                                                              |                      |
| AAATATATAT               | TTTTTGAGAA AAATATATAT                          | TTTGACATAA                                                        | CATATGCTAA                |                                                                                                                                     | GATTTGTATA AGAAATATCT TTAAAAACC      | GATTTGTATA                                                                        | 4411                 |
| AATACAGAGG               | TGCTTTTGTA                                     | TTAAGTTCGT                                                        | GAGTCGGTTT                | ATCTGTATTT                                                                                                                          | TTCGCGATTT GGAAATGCAT                | TTCGCGATTT                                                                        | 4341                 |
| TTTAAGTTTA               |                                                | CATTTTGTA                                                         | TTTGATACGT                | TTGTCGGTAC                                                                                                                          | TTATAAAAT TGAGGATGTT                 | TTATAAAAT                                                                         | 4271                 |
|                          |                                                | GCCCAAAGTG                                                        | ATAAAAGGTG                | AATATCGGTA                                                                                                                          | GTATAAGTAA                           | TTAAAAGGTG                                                                        | 4201                 |
|                          | GTCAAAAGTG                                     | AGATTACCTG                                                        | TCCARATAR AACAAGGIA       |                                                                                                                                     | GTCGGGATAG                           | TGAGAAGAGA                                                                        | 4131                 |
| TTCATATGCT               | CTCGAGGTCA                                     | ರಿತಿರಿದ್ದರು                                                       | AGCTGGGTAC                | CCCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCCC                                                                                         |                                      | GCGCAATTAA                                                                        | 4061                 |
|                          | 222222                                         | 2                                                                 | 22222                     |                                                                                                                                     |                                      |                                                                                   |                      |
|                          | XhoI                                           |                                                                   | KpnI                      |                                                                                                                                     |                                      |                                                                                   |                      |
| TACGCCAAGC               | TGACCATGAT                                     | GAAACAGCTA                                                        | TTTCACACAG                | ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC                                                        | GAATTGTGAG                           | ATGTTGTGTG                                                                        | 3991                 |
| TCCGGCTCGI               | ACTTTATGCT                                     | CAGGCTTTAC                                                        | TTAGGCACCC                | AGCTCACTCA                                                                                                                          | AATGTGAGTT                           | 3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT | <u> </u>             |

# FIG.\_32E



**SUBSTITUTE SHEET (RULE 26)** 

U

>

NGOI

Hindiii

CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG H 4 屑 AAGCTTACCA TGGCCCACGC Noti Н

E D L Y S R L V E M CGAAGACCTC TACAGCTT TAGTCGAAAT Sall GCCTCCACGC AGGGCATCTC Ø Ö H മ ರದಿಂದರಿದರಿದರು ₫ œ · A S S TOTO COLO

7

GGGAGAGAAA Ö • A T I S Q A A Y A D L C N I P S T I I K GGCCACTAIC TCCCAAGCIG CCTACGCCGA CCTGIGCAAC ATTCCGICGA CTATTAICAA BamHI 141

CAGCAAAGAA ATAATCACCG A × Ø z Н O Ø

CCTGGCGCA CTCACTGCCG CCCAGCTGTC TGCGACATAC GACAACATCC GCCTGTACAC GATACGAATC TACAACTCGA TACTAACTAC ACCCTCACGC CTTTCGACAC C H X CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT ATTTACAATT CTCAAACTGA CATTAACGGA TGGATCCTCC GCGACGACAG TIGICAAACA GCAGGTIAGC CAGTATCCGG ACTACGCGCT æ TCTTCCGTGG CACTGGTAGT GTCGAGTCGC TCGGCGCCTC 7 M 281 351 421 491

AGCCTCGAGC CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC CCGGTGGAGC Ø OF GEPRS GNOATCAGG CTTCGCGTCG TACATGAACG ATGCCTTCCA Ö p4 O NCOL 561 631

£ Z O 4 Ø ρι ĸ 3 Ħ 闰 > Ø 耳 æ Ö

| STACGC<br>G D    | CCATGGCGGT<br>E V O   | AGGGGTACGC CCATGGCGGT GTAGAGTACT | GGAGCGTTGA            | TCCTTACAGC | TCCTTACAGC GCCCAGAACA | CATTTGTCTG |
|------------------|-----------------------|----------------------------------|-----------------------|------------|-----------------------|------------|
| U                | AGTGCAG               | SCTGT                            | 3660                  | CAGGGTGTGA | ATAATGCGC             | CACGACTTAT |
| E                | 8<br>9                | CTW                              | *                     |            |                       |            |
| ၓ                | TTTGGGATGA CGAGCGGAGC | CTGTACATGG                       | TGATCAGTCA            | TTTCAGCCTC | CCCGAGTGTA            | CCAGGAAAGA |
| 8                | GGAGAGGGGG            | CCGCGTAACC                       | ACTGAAGGAT            | GAGCTGTAAA | GAAGCAGATC            | GTTCAAACAT |
| 8                | AGTTTCTTAA            | GATTGAATCC                       | TGTTGCCGGT            | CTTGCGATGA | TTATCATATA            | ATTTCTGTTG |
| AG               | AGCATGTAAT            | AATTAACATG                       | TAATGCATGA            | CGTTATTTAT | GAGATGGGTT            | TTTATGATTA |
| AT               | ATTATACATT            | Taatacgcga                       | TAGAAAACAA<br>Hindiii | AATATAGCGC | GCAAACTAGG            | ATAAATTATC |
|                  |                       |                                  | 1111111               |            | _                     |            |
|                  |                       | ClaI                             | I xbaI                | 벌          | -                     |            |
|                  |                       | 222222                           | *****                 | 2 2        |                       |            |
| ភ្ជ              | TCATCTATGT            | TACTAGATCG                       | ATAAGCTTCT            | AGAGCGGCCG | GTGGAGCTCC            | AATTCGCCCT |
| G.               | GTATTACGCG            | CGCTCACTGG                       | CCGTCGTTTT            | ACAACGTCGT | GACTGGGAAA            | ACCCTGGCGT |
| AA               | AATCGCCTTG            | CAGCACATCC                       | CCCTTTCGCC            | AGCTGGCGTA | ATAGCGAAGA            | GGCCCGCACC |
| ပ္ပ              | CCCAACAGTT            | GCGCAGCCTG                       | AATGGCGAAT            | GGGACGCGCC | CTGTAGCGGC            | GCATTAAGCG |
| 967              | GUGGLTACG             | CGCAGCGTGA                       | CCGCTACACT            | TGCCAGCGCC | CTAGCGCCCG            | CTCCTTTCGC |
| ည္သ              | CCTTTCTCG             | CCACGTTCGC                       | CGGCTTTCCC            | CGTCAAGCTC | TAAATCGGGG            | GCTCCCTTTA |
| TTA              | TTAGTGCTTT            | ACGGCACCTC                       | GACCCCAAAA            | AACTTGATTA | GGGTGATGGT            | TCACGTAGTG |
| CTG              | CTGATAGACG            | GLITITICGCC                      | CTTTGACGTT            | GGAGTCCACG | <b>TTCTTTAATA</b>     | GIGGACICIT |
| GGS              | SGAACAACAC            | TCAACCCTAT                       | CTCGGTCTAT            | TCTTTTGATT | TATAAGGGAT            | TTTGCCGALT |
| GGH              | GGTTABAAA             | TGAGCTGATT                       | TAACAAAAT             | TTAACGCGAA | TTTTAACAAA            | ATATTAACGC |
| E<br>C<br>C<br>T | GGTGGCACTT            | TTCGGGGAAA                       | TGTGCGCGGA            | ACCCCTATT  | GTTTATTTT             | CTABATACAT |
| ATC              | ATCCCCTCAT            | GAGACAATAA                       | CCCTGATAAA            | TGCTTCAATA | ATATTGAAAA            | AGGAAGAGTA |
| AC A             | ACATTTCCGT            | GICGCCCTIA                       | TTCCCLTTLT            | TGCGGCATTT | TGCCTTCCTG            | TTTTGCTCA  |
| E<br>U           | CTGGTGAAAG            | TABABGATGC                       | TGAAGATCAG            | TIGGGIGCAC | GAGTGGGTTA            | CATCGAACTG |
| ပ္ပပ္ပ           | SCGGTAAGAT            | CCTTGAGAGT                       | TTTCGCCCCG            | AAGAACGTTT | TCCAATGATG            | AGCACTTTTA |
| ATG              | ATGTGGCGCG            | GTATTATCCC                       | GTATTGACGC            | CGGGCAAGAG | CAACTCGGTC            | GCCGCATACA |
| AAT              | AATGACTTGG            | TTGAGTACTC                       | ACCAGTCACA            | GAAAAGCATC | TTACGGATGG            | CATGACAGTA |
| ູ່ປີ             | GCAGTGCTGC            | CATAACCATG                       | AGTGATAACA            | CTGCGGCCAA | CTTACTTCTG            | ACAACGATCG |
| GGZ              | GGAGCTAACC            | GCTTTTTTGC                       | ACAACATGGG            | GGATCATGTA | ACTCGCCTTG            | Atcortgoga |
| A                | AATGAAGCCA            | TACCAAACGA                       | CGAGCGTGAC            | ACCACGATGC | CTGTAGCAAT            | GGCAACAACG |
| TA C             | TATTAACTGG            | CGAACTACTT                       | ACTCTAGCTT            | CCCGGCAACA | ATTAATAGAC            | TGGATGGAGG |
| 5                | I'GCAGGACCA           | CITICITECECT                     | CGGCCCTTCC            | GGCTGGCTGG | TTTATTGCTG            | ATAAATCTGG |

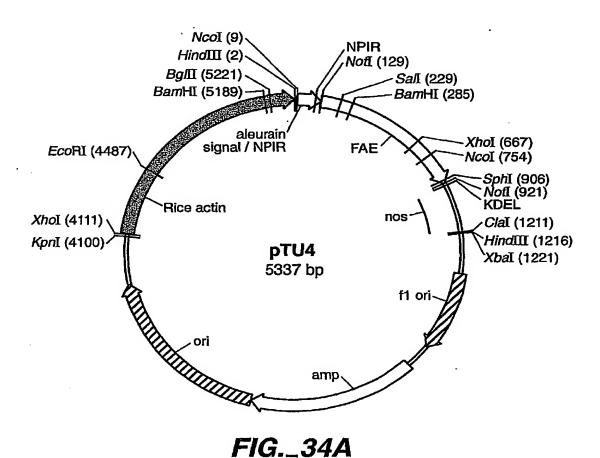
-1G.\_33C

| CCGTATCGTA<br>ATAGGTGCCT<br>TAAAA       | TTAACGTGAG | TTTTTCTGC<br>ATCAAGAGCT  | TCTAGTGTAG | ATCCTGTTAC<br>TACCGGATAA | CTACACCGAA | GACAGGTATC | GGTATCTTTA | GGGCCGGAGC | GCTCACATGT | ATACCGCTCG   | AUGUAAAUUG                                       | AAAGCGGGCA |            | CCATGATTAC | ĭ    | 2 2 3  | GAGGTCATTC        | AAAAGTGAAA | TTTACTCTTT | ATTGGTTTTT | TTTTGTAAT  | <b>TTGAGAAAA</b>    |       | GCAGCGATGG | CTABAGICCI | CCACCCCAGT | CGCACGTCTC | GrcGraaaaa | GARACGCCCC |
|-----------------------------------------|------------|--------------------------|------------|--------------------------|------------|------------|------------|------------|------------|--------------|--------------------------------------------------|------------|------------|------------|------|--------|-------------------|------------|------------|------------|------------|---------------------|-------|------------|------------|------------|------------|------------|------------|
| GTAAGCCCTC<br>GATCGCTGAG                | CCAAAATCCC | TTGAGATCCT<br>TGTTTGCCGG | ATACTGTCCT | CGCTCTGCTA               | AGCGAACGAC | GAGAAAGGCG | GGAAACGCCT | GCTCGTCAGG | CIGGCCTTTT | GAGTGGGTTG   | AGCGCCCAAT                                       | TCCCGACTGG | GCTTTACACT | ACAGCTATGA | Xbol | 1 2 2  | <b>BCCCCCCCTC</b> | TTACCTGGTC | CAAAGTGAAA | TTTTGTATGA | AGTTCGTTGC | GACATAATTT          |       | TGCCCCCGTT | AAAACAACCC | CCAACCCAAC | TCCGCACCAC | GTGGGTCCGG | CGCTTCCAAA |
| GGGCCAGATG<br>GAAATAGACA<br>ATATATACTT  | AATCTCATGA | AAGGATCTTC<br>AGCGGTGGTT | CAGATACCAA | CTACATACCT<br>GTTGGACTCA | CCCAGCTTGG | TTCCCGAAGG | GCTTCCAGGG | TTTTTGTGAT | TGGCCTTTG  | TACCECCITT   | GAMGCGGAAG                                       | ACGACAGGTT | GGCACCCCAG | CACACAGGAA | Koni | 222222 | TGGGTACCGG        | AAAGGTAAGA | AAAGGTGGCC | GATACGTCAT | TCGGTTTTTA | ATGCTAATTT          |       | TAAAATAGCT | AACATTTACA | CCAACCCAAC | CCGTGAGTTG | AAAACAGCAG | GCCCTCCCTC |
| TGCAGCACTG<br>ATGGATGAAC<br>AAGTTTACTC  | CCTTTTTGAT | GAAGAGATCA               | CAGCAGAGCG | GTCTTACCGG               | GTGCACACAG | AGCGCCACGC | GCACGAGGGA | TGAGCGTCGA | TTACGGTTCC | ATARCCGTAT   | MOTEMBLEAGE<br>MORT GRACIERS                     | TGCAGCTGGC | TCACTCALTA | ATAACAATTT |      |        | GAACAAAAGC        | AAAATAAAAC | ATCGGTAATA | TCGGTACTTT | TGTATTTGAG | AAAAACCCAT          | •     | Ataataagat | TTAGACTCAA | CAAGCCCAGC | GGCACTATCA | AAGAAAAGA  | Accadecced |
| GCGGTATCAT<br>TCAGGCAACT<br>CTGTCAGACC  | AGGTGAAGAT | AGACCCCGTA               | TAACTGGCTT | GATAAGTCGT               | CGGGGGGTTC | GCTATGAGAA | ACAGGAGAGC | ACCTCTGACT | CGCGGCCTTT | GALTCIGLGG   | GCAGCGAGTC                                       | GATTCATTAA | GTGAGTTAGC | Trereaces  |      |        | TCACTAAAGG        | GGGATAGTCC | TAAGTAAAT  | GGATGTTTTG | AATGCATATC | AATATCTTTA          |       | ACAATGAACA | AAAGATAAAC | CGATCCATAG | CICCACCCCC | AAAGAAAAA  | GCGAGCAGCG |
| COTGGGGTCTC<br>COACGGGGAG<br>GCATTGGTAA | AAAAGGATCT | ACTERECETC               | TTTCCGAAGG | TGCCAGTGGC               | TCGGGCTGAA | TACAGCGTGA | CAGGGTCGGA | GGGTTTCGCC | ACCCCACCAA | GI-TAICCCCT. | ACCOUNT OF THE CO                                |            | CGCAATTAAT | Treteregra |      |        | CAATTAACCC        | GAAGAGAGTC | AAAGGTGGTA | TAAAAATTGA | GCGATTTGGA | TTGTATAAGA<br>ECORI | 22222 | GGCGAATTCC | TAGTAAAATA | GTGCTATGCA | GGCAAATAGT | AAAAAAAAG  | GAGGAGGATC |
| AGCCGGTGAG<br>GTTATCTACA<br>CACTGATTAA  | TTTTAATTT  | TITICGITC<br>GCGTAATCIG  | ACCAACTCTT | CAGTGGCTGC               | GGCGCAGCGG | CTGAGATACC | CGGTAAGCGG | TAGTCCTGTC | CTATGGAAAA |              | を できることを できる |            | GTGAGCGCAA | GCTCGTATG  |      |        | GCCAAGCGCG        | ATATGCTTGA | ACATCAGTTA | TCTACTATTA | AAGTTTATTC | Acagaggat           |       | TATATATTCA | GTATTTTTC  | AAAGCCCAAA | GCAGCCAACT | GCAGCCAAAA | CCGGAAAAGC |
| 2731<br>2801<br>2871                    | 2941       | 3081                     | 3151       | 1657<br>3201<br>3201     | 3361       | 3431       | 3501       | 3571       | 3641       | 2707         | 70/07                                            | 1000       | 1000       | 3991       |      |        | 4061              | 4131       | 4201       | 4271       | 4341       | 4411                |       | 4481       | 4551       | 4621       | 4691       | 4761       | 4831       |

-1G.\_33D

| A U                                                                                            | Ф                                                                           | Ø                                                                            |       |        | Ę+                                                                                | ø                                                                                 |                                     |
|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------|-------|--------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-------------------------------------|
| ACCACCACC<br>GTAACCACC                                                                         | GGTAGTTTG                                                                   | GGCGTCTCC                                                                    |       |        | TTCTTCTT                                                                          | GACAAATGC                                                                         |                                     |
| TACCACCACC                                                                                     | CTTTGGCCTT                                                                  | TCTCGCGGCT                                                                   | н     | 2 2    | TCTTCTTCT                                                                         | CATGATTTGT                                                                        |                                     |
| CCCCCAACCC                                                                                     | CGGTCTCGAT                                                                  | AGGGGCGGGA                                                                   | Bglii | 22222  | CGGATGTAGA                                                                        | TTTTTCTTT                                                                         |                                     |
| TCCTCCCATC                                                                                     | TTTTCGICT                                                                   | GGTGCGCGGG                                                                   |       |        | ATGGGGCTCT                                                                        | TCATCGGTAG                                                                        |                                     |
| CCCCCCCTC                                                                                      | CTCCGTTTTT                                                                  | CGCCCAGATC                                                                   |       | 3      | CTCGCGGGGA                                                                        | TCAGCATTGT                                                                        |                                     |
| TATATACATA                                                                                     | CGCCCCTCTC CTCTTTCTTT CTCCGTTTTT TTTTCGTCT CGGTCTCGAT CTTTGGCCTT GGTAGTTTGG | GCGGCTTCGT                                                                   | BamHI | ****** | GGCCCGGATC                                                                        | TTTGAATCCC                                                                        |                                     |
| CCATCGCCAC TATATACATA CCCCCCTC TCCTCCCATC CCCCCAACCC TACCACCACC ACCACCACCA CCTCCTCCCC CTCCCTCC | これつこうこうじょう                                                                  | GTGGGCGAGA GCGGCTTCGT CGCCCAGATC GGTGCGCGGG AGGGGCGGGA TCTCGCGGCT GGCGTCTCCG |       |        | 5181 GGCGTGAGTC GGCCCGGATC CTCGCGGGGA ATGGGGCTCT CGGATGTAGA TCTTCTTTCT TTCTTCTTTT | 5251 TGTGGTAGAA TTTGAATCCC TCAGCATTGT TCATCGGTAG TTTTTCTTTT CATGATTTGT GACAAAFGCA | くりかけで 日本日本のはないという しゅうしゅうしゅうしょう すってい |
|                                                                                                | 5041                                                                        | 5111                                                                         |       |        | 5181                                                                              | 5251                                                                              | F221                                |

## FIG.\_33E



**SUBSTITUTE SHEET (RULE 26)** 

MGOI

HindIII

Н

AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG H 哞

GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGC Ø GCCCGTCACC ρŧ 7 T p, CICCICCIIC GCCGACICCA Ø 4 SEU Ø · A B S TCGCCTCCTC o •

141

71

D L C N I P S T I I K G E K I Y N S Q T D I N GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAACT GACATTAACG BemHI 211

GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA GTGATACGAA CCGCCTGTAC ACCTTCGGCG AACCGCGAG CGGCAATCAG OCCAGIATCC GGACTACGCG CIGACCGIGA CCGGCCACKC CCTCGGCGCC ICCCTGGCGG CACTCACTGC TCTACAACTC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA CACGGIGGAT ATTATATING ATGGGICTC GICCAGGACC AAGICGAGIC GCTIGICAAA CAGCAGGITA **Z** 四日 **₹** A Q U Q V X L P F D CGCCCAGCTG TCTGCGACAT ACGACAACAT 2 SAR \ 0 0 9 н • 281 351 421 491

GACGCAGTAT TTCCGGGTCA **S** MCOI GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC A μ Ø ໝ ď Ω Z z Þ מ Þ 4 631

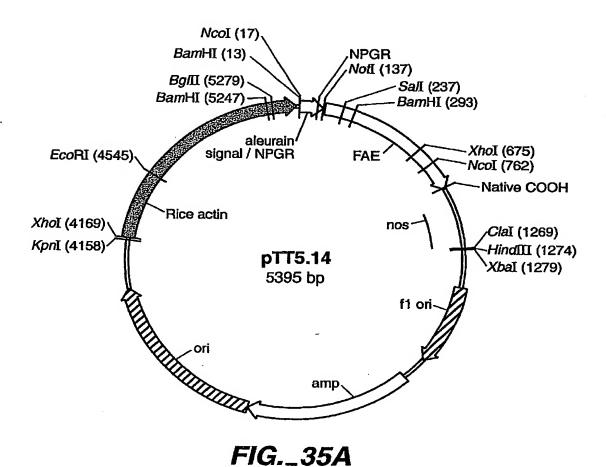
M > O Ħ ď þı O Q ρι Н Z ρι Н U A Z 4 耳

FIG.\_34B

|                                                                 | •                                        |                                                                                                       |                     | •                                                                                                                                                                                                                                                                                                            |
|-----------------------------------------------------------------|------------------------------------------|-------------------------------------------------------------------------------------------------------|---------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GTGTAGAGTA<br>C C E<br>GTGCTGTGAG<br>Spbl                       | A C T W<br>GCATGCACCT                    | atttggcaat<br>Tgaattacgt<br>Tagagtcccg<br>Tcgcgcgcgg                                                  |                     | CTATAGTGAG<br>GTTACCCAAC<br>CGGGCGGGT<br>GCGTTCTTCC<br>TAGGGTTCCG<br>TTGTTCCAAA<br>TTTCGACATT<br>TTTCAAATTT<br>TATCAAATAT<br>TATCAAATAT<br>TATCAAATAT<br>TATCAAATTT<br>TATCAAATAT<br>TATCAAATAT<br>TAAAGTTCTGA                                                                                               |
| GCAGGGGTAC GCCCATGGCG<br>C T G D E V Q<br>TGCACTGGGG ATGAAGTGCA | T S G<br>Gacgagcggc                      | TCGTTCAAAC<br>TAATTTCTGT<br>TTTTTATGAT<br>GGATAAATTA                                                  |                     | CCAATTCGCC<br>AAACCCTGGC<br>GCGCATTAAG<br>GGCTCCTTT<br>GTTCACGTAG<br>ATTTTGCCGA<br>AAATATTAAC<br>TACAAAAAA<br>TGCTAAAC<br>TGCCAAAAAA<br>TGCCGCATA<br>TGCCGCATA                                                                                                                                               |
| GCAGGGGTAC: (<br>C T G D<br>TGCACTGGGG ;                        | r f g m<br>Attitgggat                    | aagaagcaga<br>gattatcata<br>atgagatggg<br>gcgcaaacta                                                  |                     | CGGTGGAGCT<br>GTGACTGGGA<br>TATATAGCGC<br>CCCTAGCGCC<br>TCTAAATCGG<br>TCTTATAAA<br>TTTATAAGG<br>AATTTTAAA<br>TTGCCTTCC<br>TAATATTTT<br>TATATTTTTTTTTT                                                                                                                                                        |
| CCCCGGTGGA<br>T F V<br>CACATTTGTC                               | H T T Y F G M<br>CACACGACTT ATTTTGGGAT   | PLKDELL CCACTGAAGG ATGAGCTGTA CCTGTTGCCC GTCTTGCGAT TGTAATGCAT GACGTTATTT GATAGAAAC AAAATATAGC HIMILI | ADGL                | CTAGAGCGGC<br>CCAGCTGGCG<br>ATGGCACGCG<br>CCCGTCAAGC<br>AAACTTGAT<br>TTGGAGTCCA<br>ATTTAACGCG<br>GAACCCCTAT<br>ATTGCGCCAT<br>AATGCTTCAA<br>TTTGCGCCAT<br>AGTTGGGTGC<br>CGAAGAACGT<br>CCGAAGAACGT<br>CCGAAGAACGT                                                                                              |
| CCAAACCTGC<br>A Q N<br>GCGCCCAGAA                               | n n a<br>gaataatgeg                      | A CCACTGAAGG TA CCTGTTGCCG TA TGTAATGCAT C GATAGAAAAC Hindili                                         |                     | GARGET<br>GETCGET<br>CITTCC<br>GCCTAC<br>GCTAC<br>GCTAC<br>CCCCAAA<br>ACAAA<br>ACAAA<br>ACAAA<br>TCGCC<br>TCGCT<br>TCGCC<br>TCGCT<br>TCGCC<br>TCGCT                                                                                                                                                          |
| CGACGGCATC CCAAACCTGC<br>D P Y S A Q N<br>GATCCTTACA GCGCCCAGAA |                                          | GGA<br>GGA<br>AAC<br>ACG                                                                              | ול לבו<br>יינים לבו | GTTACTAGAT<br>CGCGCCACT<br>TGCGCACGT<br>CGCCCACGTT<br>TTACGGCGC<br>TTACGGCGC<br>TTACGGCGA<br>ACTCAACCCT<br>ACTCAACCCT<br>ACTCAACCCT<br>ACTCAACCCT<br>ACTCAACCCT<br>ACTCAACAACAAT<br>ATCCTTGAGA<br>ATCCTTGAGA<br>ACCCTTGAGA<br>ACCCTTGAGA<br>ACCCTTGAGA<br>ACCCTTGAGACCAACCCT                                 |
| CTCATGCCAA • W S V CTGGAGCGTT                                   | A Q G G Q G V GCCCAGGGCG GACAGGGTGT NotI | GGCCGGTCGC AAAGTTTCTT TAAGCATGTA CAATTATACA                                                           |                     | TGTCATCTAT<br>TCGTATTACG<br>TTAATCGCCT<br>TTCCCAACAG<br>GTGGTGGTTA<br>ATTTAGTGCT<br>CCCTGATTACA<br>TTGGTTAAAA<br>TTGGTTAAAA<br>TTGGTTAAAA<br>TTGGTTAAAA<br>TAGGTGGCAC<br>GTATCCGCTC<br>CAACATTTCC<br>CAACATTTCC<br>CAACATTTCC<br>CAACATTTCC<br>CAACATTTCC<br>CAACATTTCC<br>CAACATGTGGCAAAG<br>CTATGTGGCGTAAG |
| 701                                                             | 841                                      | 911<br>981<br>1051<br>1121                                                                            |                     | 1120<br>1120<br>1120<br>1120<br>1120<br>1120<br>1120<br>1120                                                                                                                                                                                                                                                 |

| GAACCGGAGC<br>CGTTGCGCAA<br>GGCGGATAAA<br>GGAGCCGGTG<br>TAGTTATCTA               | CICACIGALIT<br>CATTITIAAT<br>AGTITICGIT<br>GCGCGTAATC<br>CTACCAACTC<br>AGCCGTAGIT   | AAGGCGCAGC<br>AACTGAGATA<br>TCCGGTAAGC<br>TATAGTCCTG<br>GCCTATGGAA<br>GTTCTTTCCT               | CGCCTCTCCC<br>CAGTGAGCGC<br>CCGGCTCGTA<br>ACGCCAAGCG         | TCATATGCTT<br>AAACATCAGT<br>TTTCTACTAT<br>TTAGGTTTAT<br>ATACAGAGGG               | gggtatttt<br>Ctaaagccca  |
|----------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|--------------------------------------------------------------|----------------------------------------------------------------------------------|--------------------------|
| TGATCGTTGG ATGGCACCAA ACTGGATGGA TGATAAATCT TCCCGTATCG                           | TTTAAAACTT<br>CCTTAAACGTG<br>CCTTATTTTCT<br>GGATCAAGAG<br>CTTCTAGTGT<br>TAATCCTGGTT | GTTACCGGAT<br>ACCTACACGG<br>CGGACAGGTA<br>CTGGTATCTT<br>GGGGGGCGGA<br>TTGCTCACAT               | ATACGCAAAC<br>GGAAAGCGGG<br>CTTTATGCTT<br>GACCATGATT<br>XhoI | TCGAGGTCAT<br>TCAAAAGTGA<br>AATTTACTCT<br>GAATTGGTTT<br>GCTTTTGAGAAA             | TTGCAGCGAT<br>CCCTAAAGTC |
| TAACTCGCCT<br>GCCTGTAGCA<br>CAATTAATAG<br>GGTTTATTGC<br>TGGTAAGCCC               | TTTAGATTGA<br>GACCAAAATC<br>TCTTGAGATC<br>TTTGTTTGCC<br>AAATACTGTC<br>CTCGCTCTGC    | CAAGACGATA<br>GGAGCGAACG<br>GGGAGAAGG<br>GGGGAAACGC<br>ATCCTCGTCA<br>TGCTGGCCTT<br>TTGAGTGAGC  |                                                              | GGGCCCCCC<br>GATTACCTGG<br>CCCAAAGTGA<br>ATTTTTGTAT<br>TAAGTTCGTT<br>TTGACATAAT  | CTTGCCCCCG               |
| GGGGATCATG<br>ACACCACGAT<br>TTCCCGGCAA<br>CCGGCTGGCT<br>TGGGGCCAGA               | TCATATATAC<br>ATAATCTCAT<br>CAAAGGATCT<br>CCAGCGGTGG<br>GGCAGATACC<br>GCCTACATAC    | GGGTTGGACT<br>AGCCCAGCTT<br>GCTTCCCGAA<br>GAGCTTCCAG<br>GATTTTTGTG<br>CCTGGCCTTT               | AGGAAGCGGA<br>GCACGACAGG<br>TAGGCACCCC<br>TTCACACAGG         | GCTGGGTACC<br>ACAAAGGTAA<br>TAAAAGGTGG<br>TTGATACGTC<br>AGTCGGTTTT<br>ATATGCTAAT | attaaaatag<br>aaaacattta |
| GCACAACATG GACGAGCGTG TTACTCTAGC CTCGGCCCTT ATTGCAGCAC                           | CCAAGTTTAC<br>ATCCTTTTTG<br>TAGAAAGAT<br>ACCACCGCTA<br>TTCAGCAGAG                   | GTGTCTTACC<br>TCGTGCACAC<br>AAAGCGCCAC<br>GCGCACGAGG<br>CTTGAGCGTC<br>TTTTACGGTT<br>GGATAACGGT | TCAGTGAGCG<br>AATGCAGCTG<br>GCTCACTCAT<br>GGATAACAAT         | GGGAACAAA<br>CCAAAATAAA<br>ATATCGGTAA<br>TGTCGGTACT<br>TCTGTATTTG<br>TAAAAAACCC  | Caataataag<br>acttagactc |
| CCGCTTTTTT<br>CATACCAAAC<br>GGCGAACTAC<br>CACTTCTGCG<br>TCGCGGTATC               | AACTGTCAGA<br>CTAGGTGAAG<br>TCAGACCCG<br>AAACAAAAA<br>GGTAACTGGC<br>TTCAAGAACT      | GCGATAAGTC<br>AACGGGGGGT<br>GAGCTATGAG<br>GAACAGGAGA<br>CCACCTCTGA<br>AACGCGGCCT<br>CTGATTCTGT | GCGCAGCGAG<br>CCGATTCATT<br>ATGTGAGTTA<br>AATTGTGAGC         |                                                                                  | Ccacaatgaa<br>Taaaagataa |
| AAGGAGCTAA<br>TGAATGAAGC<br>ACTATTAACT<br>GTTGCAGGAC<br>AGCGTGGGTC<br>CACGACGGGG | AAGCATTGGT<br>TTAAAAGGAT<br>CCACTGAGCG<br>TGCTGCTTGC<br>TTTTTCCGAA                  | GCTGCCAGTG<br>GGTCGGGCTG<br>CCTACAGCGT<br>GGCAGGGTCG<br>TCGGGTTTCG<br>AAACGCCAGC               | GAACGACCGA<br>CGCGCGTTGG<br>AACGCAATTA<br>TGTTGTGTGG         | CGCAATTAAC GAGAAGAGAG TAAAAGGTGG TATAAAAATT TCGCGATTTG ATTTGTATAA                | Caggcgaalt<br>Tctagtaaaa |
| 2451<br>2591<br>2591<br>2731<br>2801                                             | 2871<br>2941<br>3011<br>3081<br>3151                                                | 3291<br>3361<br>3501<br>35701<br>3711                                                          | 3481<br>39851<br>3981<br>3981                                | 44444<br>001044<br>001084<br>10001<br>11111                                      | 4481<br>4551             |

## FIG.\_34E



SUBSTITUTE SHEET (RULE 26)

MCOI BanHI

~~~~~

CCIGGCGCIC GCCGIGCIGG CCACGGCCGC Þ 4 CCTGACGCCG AGGATCCATG GCCCACGCCC GCGTCCTCCT H > 畔 ď, 耳 4 Ħ

GCCTCCTCCT CCTCCTTCGC CGACTCCAAC CCGGGCCGGC CCGTCACCGA CCGCGGGC A Ę٠ > щ p4 O ρι Z Ø А ¢ [± Ø Ø 93 Ø 4 CGTCGCCGTC > ¥ >. Noti 7

A S T Q G I S E D L Y S R L V E M A T I S Q A GCCRCTACCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT GGCCACTATC TCCCAAGCTG SalI 141

CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA GGGAGAGAAA ATTTACAATT CTCAAACTGA O! N N H 瓦瓦 U × н н Ø ρį r G Bamer A 4 þ 211

D T N L Q L D T N Y T L T P F D T L P Q C N G GATACGAATC TACAACTAC ACCTCACGC CTTTCGACAC CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA GTCGAGTCGC TTGTCAAACA · I N G W I L R D D S S K E I I T V F R G T G S CATTAACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATAATCACCG TCTTCCGTGG CACTGGTAGT α α α 0 W V S V X O C H > H 281 351 421

GCAGGTTAGC CAGTATCCGG ACTACGCGCT GACCGTGACC GGCCACKCCC TCGGCGCCTC CCTGGCGGCA LTAA QLUSATYDNIRLYTFGEPRS G. H. K. L. G. A. S. T A T YAL A F F > 0 491 561

F A S Y M N D A F Q A S S P D T T Q Y F CTTCGCGTCG TACATGAACG ATGCCTTCCA AGCCTCGAGC CCAGATACGA CGCAGATTT NGOI • N Q A GCAATCAGGC 631

O U Ħ O O! F4 Þ Д p, Н Z ρι н Ü A Z 4 × H Þ æ

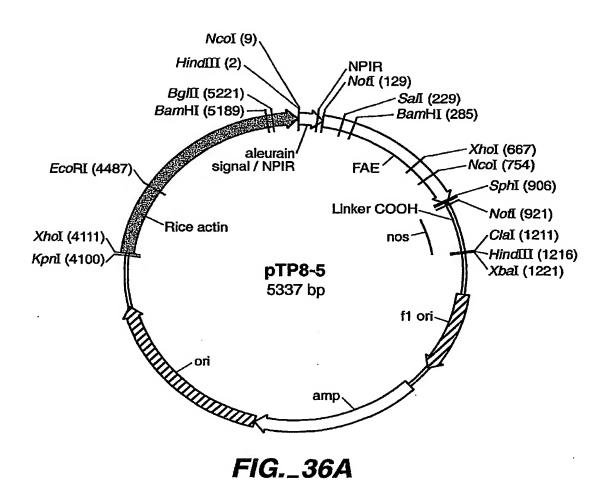
ַ ט ַ	•									_	_			_				_					_						
CCATGGCGGT E V Q GAAGTGCAGT	COAGCGGAGC	GGAGAGGGGG	AGCATGTAAT	ALTATACALL	TCATCTATGT				GTATTACGCG	AATCGCCTTG	CCCAACAGTT	GGTGGTTACG	TCCLTTCTCG	TTAGTGCTTT	CTGATAGACG	GGAACAACAC	GGTTAAAAA	GGTGGCACTT	ATCCGCTCAT	ACATITICCGI	CTGGTGAAAG	GCGGTAAGAT	ATGTGGCGCG	AATGACTTGG	GCAGTGCTGC	GGAGCTAACC	AATGAAGCCA	TATTAACTGG	
¥ - E:	f g m t b g a Tygggatga cgagcggagc	TOGATGICCT	AATTACGTTA	GAGTCCCGCA	есасасадие				ATAGTGAGTC	TACCCAACTT	GATCGCCCTT	CGGCGGGTGT	TTTCTTCCT	GGGTTCCGAT	GGCCATCGCC	GTTCCAAACT	TCGGCCTATT	TTACAATTTA	TCAAATATGT	TGAGTATTCA	CCCAGAAACG	GATCTCAACA	AAGTTCTGCT	CTATTCTCAG	AGAGAATTAT	GAGGACCGAA	ACCEGAGCTG	TTGCGCAAAC	
	CACGACTTAT	CCAGGAAAGA	ATTTCTGTTG	TTTATGATTA	Ataaattatc				AATTCGCCCT	ACCCTGGCGT	CCCCCCACC	GCATTAAGCG	CTCCTTTCGC	GCTCCCTTTA	TCACGTAGTG	GTGGACTCTT	TTTGCCGATT	ATATTAACGC	CTAAATACAT	AGGAAGAGTA	TTTTTGCTCA	CATCGAACTG	AGCACTTTTA	GCCGCATACA	CATGACAGTA	ACAACGATCG	ATCGTTGGGA	GGCAACAACG	
A Q N T SCCCAGAACA C	ATAATGCGCA	CCCGAGTGTA	-	GAGATGGGTT	GCAACTAGG				GTGGAGCTCC	GACTGGGAAA	ATAGCGAAGA	CTGTAGCGGC	CTAGCGCCCG	TAAATCGGGG	GGGTGATGGT	TTCTTTAATA	TATAAGGGAT	TTTTAACAAA	<b>GITTLAITIT</b>	ATATTGAAAA	Teccificate	GAGTGGGTTA	TCCAATGATG	CAACTCGGTC	TTACGGATGG	CTTACTTCTG	ACTCGCCTTG	CTGTAGCAAT	
ACGGCATCCC AAACCTGCCC CCGGTGGAGC P Y S A Q N T F V C TCCTTACAGC GCCCAGAACA CATFTGTCTG	CAGGGTGTGA ATANTGCGCA CACGACTTAT	TTTCAGCCTC (		_	AATATAGCGC (		<b></b>	1 1 1	AGAGCGGCCG (	ACAACGTCGT (	AGCTGGCGTA	GGGACGCGCC	TGCCAGCGCC	CGTCAAGCTC	AACTTGATTA	GGAGTCCACG '	TCTTTTGATT '	TTAACGCGAA '	ACCCCTATTT	TGCTTCAATA	TGCGGCATTT	TTGGGTGCAC	AAGAACGTTT	CGGGCAAGAG	GAAAAGCATC	CTGCGGCCAA	GGATCATGTA	ACCACGATGC	
	CCAGGGCGGA (		TOTTGCCGGT (	_	TAGAAAACAA 1 Hindiii	2 2 2 2 2 2	XbaI	*******	ATAGCTTCT ;	CCGTCGTTTT	CCCTTTCGCC	AATGGCGAAT	CCGCTACACT	CGGCTTTCCC	GACCCCAAAA	CTTTGACGTT	CTCGGTCTAT	TAACAAAAT	rerececeda	CCCTGATAAA	TTCCCTTTTT	TGAAGATCAG	TTTCGCCCCG	GTATTGACGC	ACCAGTCACA	AGTGATAACA	ACAACATGGG	CGAGCGTGAC	
CCGGGTCACT CATGCCAACG V E Y W S V D GTAGAGTACT GGAGCGTTGA	CTGTGAGGC (		GEGEGIRACE A	AATTAACATG	TAATACGCGA 1		ClaI	222222	TACTAGATCG 1	CGCTCACTGG (	CAGCACATCC (	_	CGCAGCGTGA (	CCACGTTCGC (	ACGCCACCTC (	GTTTTTCGCC (	CAACCCTAT (	TGAGCTGATT :	TTCGGGGAAA	GAGACAATAA (	GTCGCCCTTA !	-	CCTTGAGAGT	GTATTATCCC (	TTGAGTACTC 1			TACCAPACGA	
701 0	841 0		1051		-				1261 1		_	_	_	•	~			1891	1961	2031 (	2101	_	~	2311 (	2381	2451 (	2521 (	2591	

CGTGGGTCTC	GCATTGGTAA	AAAAGGATCT	ACTGAGCGTC	CTGCTTGCAA	TTTCCGAAGG	GCCACCACTT	TGCCAGTGGC	TCGGGCTGAA	TACAGCGTGA	CAGGGTCGGA	GGGTTTCGCC	ACGCCAGCAA	GTTATCCCCT	ACGACCGAGC	CGCGTTGGCC	CGCAATTAAT	TTGTGTGGAA	CAATTAACCC			GAAGAGAGTC	AAAGGTGGTA	TAAAAATTGA	GCGATTTGGA	TTGTATAGA	ECORI	2	GGCGAATTCC	TAGTAAATA	GTGCTATGCA	GGCAAATAGT	aaaaaaag Gaggaggatc
AGCCGGTGAG	CACTGATTAA	TTTTTTT	TITICGITICC	GCGTAATCTG	ACCAACTCTT	CCGTAGTTAG	CAGTGGCTGC	GGCGCAGCGG	CTGAGATACC	CGGTAAGCGG	TAGTCCTGTC	CTATGGAAAA	TCTTTCCTGC	CCGCAGCCGA	CCTCTCCCCG	GTGAGCGCAA	GGCTCGTATG	GCCAAGCGCG			ATATGCTTGA	ACATCAGTTA	TCTACTATTA	AAGTTTATTC	ACAGAGGGAT		•	TATATATTCA	GTATTTTTC	AAAGCCCAAA	GCAGCCAACT	CCGGAAAAGC
ATAAATCTGG	ATAGGTGCCT	TAAAACTTCA	TTAACGTGAG	THITHCIGC	ATCAAGAGCT	TCTAGTGTAG	ATCCTGTTAC	TACCGGATAA	CTACACCGAA	GACAGGTATC	GGTATCTTTA	GGGGCGGAGC	GCTCACATGT	ATACCGCTCG	ACGCAAACCG	APAGCGGGCA	TTATGCTTCC	CCATGATTAC	Ä	1	GAGGTCALTC	AAAAGTGAAA	TTTACTCTTT	ATTGGTTTT	TTTTGTAAAT			TTGAGAAAA	GCAGCGATGG	CTAAAGTCCT	CCACCCAGT	GEGEGGGGG
TTTATTGCTG	GATCGCTGAG	TAGATTGALT	CCAAAATCCC	TTGAGATCCT	TGTTTGCCGG	ATACTGTCCT	CGCTCTGCTA	AGACGATAGT	AGCGAACGAC	GAGAAAGGCG	GGAAACGCCT	GCTCGTCAGG	CTGGCCTTTT	GAGTGAGCTG	AGCGCCCAAT	TCCCGACTGG	GCTTTACACT	ACAGCTATGA	XhoI	222	ರ್ವವವವವವಾಶ	TTACCTGGTC	CAAAGTGAAA	TTTTGTATGA	AGTICGIIGC			GACATAATTT	Teccccent	AAACAACCC	CCAACCCAAC	TCCGCACCAC
GGCTGGCTGG	GAAATAGACA	ATATATACTT	AATCTCATGA	AAGGATCTTC	AGCGGTGGTT	CAGATACCAA	CTACATACCT	GTTGGACTCA	CCCAGCTTGG	TTCCCGAAGG	GCTTCCAGGG	<b>TTTTTGTGAT</b>	TGGCCTTTTG	TACCGCCTTT	GAAGCGGAAG	ACGACAGGTT	GGCACCCCAG	CACACAGGAA	KpnI	******	TGGGTACCGG	AAAGGTAAGA	AAAGGTGGCC	GATACGTCAT	TCGGTTTTTA			ATGCTAATTT	TARARTAGCT	AACATTTACA	CCAACCCAAC	CCGTGAGTTG AAAACAGCAG
CGGCCCTTCC	ATGGATGAAC	AAGTTTACTC	CCTTTTTGAT	GALABGATCA	CACCGCTACC	CAGCAGAGCG	GTAGCACCGC	Grcttacce	GTGCACACAG	AGCGCCACGC	GCACGAGGGA	TGAGCGTCGA	TTACGGTTCC	ATAACCGTAT	AGTGAGCGAG	TGCAGCTGGC	TCACTCATTA	ATAACAATTT		- *	GAACAAAAGC	AAAATAAAAC	ATCGGTAATA	TCGGTACTT	TGTATTTGAG	-		AAAACCCAT	ATAATAAGAT	TTAGACTCAA	CAAGCCCAGC	AAGAAAAAGA
CTTCTGCGCT	TCAGGCAACT	CTGTCAGACC	AGGTGAAGAT	AGACCCCGTA	ACAAAAAAC	TAACTGGCTT	CAAGAACTCT	GATAAGTCGT	CGGGGGGTTC	GCTATGAGAA	ACAGGAGAGC	ACCTCTGACT	CGCGGCCTTT	GATTCTGTGG	GCAGCGAGTC	GATTCATTAA	GTGAGTTAGC	TTGTGAGCGG			TCACTAAAGG	GGGATAGTCC	TAAGTAAAT	GGATGTTTTG	AATGCATATC			AATATCTTTA	ACAATGAACA	AAAGATAAAC	CGATCCATAG	AAAGAAAAAA
2731	_	2941	3011	3081	3151	3221	3291	3361	3431	3501	3571	3641	3711	3781	3851	3921	3991	4061			4131	4201	4271	4341	4411		,	4481	4551	4621	4697	4831

FIG.\_35D

TATATACATA CCTCGCTGCC CTCTTTCTTT GCGGCTTCGT	Bamhi	tttgaarccc gagctttttt
CCATCGCCAC CCTCCTCCC CGCCCTCTC GTGGGCGAGA	GGCGTGAGTC	TGTGGTAGAA GCCTCGTGCG
GAAACGCCCC ACCACCACCA GTAACCACCC GGTAGTTTGG	GGCGTCTCCG	ttcttcttt Gacaaatgca
CGCTTCCAAA TACCACCACC GCCGCCGCCG	TCTCGCGGCT II	TCTTCTTTCT CATGATTTGT
GCCCTCCTC CCCCCAACCC CTCCCCCTCC CGGTCTCGAT	AGGGGCGGGA TV Bglii	CGGATGTAGA TCT TTTTTCTTTT CAT
GCGAGCAGCG ACGAGGCCCG GCCTCCCTC CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTT TCCTCCCATC CCCCCAACC TACCACCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCCCTCC GCCGCCGCG GTAACCACC CGCCCTCTC CTCTTTCTTT CTCCGTTTTT TTTTCGTCT CGGTCTCGAT CTTTGGCCTT GGTAGTTTGG GTGGGCGAGA GCGGCTTCGT	GGTGCGCGG	atgegectet Tcategetag
GCGAGCAGCG ACGAGGCCCG GCCTCCCTC CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCCAACCC TACCACCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCCCCTCC GCCGCCGCG GTAACCACCC CGCCCTCTC CTCTTTTTTTTTT	Bamhi CGCCCAGAIC GGIGCGGGG AGGGGCGGGA ICICGCGGCI GGCGICICCG GGCGIGAGIC GGCCCGGAIC Bamhi	CTCGCGGGGA ATGGGGCTCT CGGATGTAGA TCTTCTTTCT TTCTTCTTTT TGTGGTAGAA TTTGAATCCC TCACCAGTGT TCACCATTGT GACAATGCA GCCTCGTGCG GAGCTTTTTT GTAGCA GCCTCGTGCG GAGCTTTTTT GTAGCA
4901 4971 5041 5111	5181	5251 5321 5391

### FIG.\_35E



SUBSTITUTE SHEET (RULE 26)

Ncol

Hindiii

GGCCACGGCC GCCGTCGCCG H 4 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT Н H Þ 앩

1

GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC Ø GCCCGTCACC H Z S S F A D S N P I R CTCCTCCTTC GCCGACCG r 4 þ H 四四 TCGCCTCTC o •

Sall

141

Ø GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAAACT GACATTAACG A H O 10 Z н × 闰 Ø Н H Ø Z rı U BamHI 211

GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA GTGATACGAA TCTACAACTC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA CACGGIGGAT AFTATATIGG ATGGGICTCC GICCAGGACC AAGTCGAGTC GCFFGTCAAA CAGCAGGTTA r F D **z** æ Þ: Þ A O O O O A × H U PFD H A 又 W V 33 so co R C C ¥ © © M H W 281 421 351

CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC 64 64 13 13 7 1 æ B × H GCCAGTATCC 491 561

XhoI

GACGCAGTAT TTCCGGGTCA MOOI Ø A F A S Y M N D A F Q A S B P D T GCCTTGGGGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC Д 631

× ഠ O H ď ы O O) ρι ρą н Z Q, н Ø А Z d щ

-1G.\_36B

TAAAGTTCTG GTGCTGTGAG TAGAGTCCCG TCGCGCGCGG GCTTTCTTCC TAGGGTTCCG attcaaatat TATGAGTATT CACCCAGAAA TGGATCTCAA CACTATTCTC CCCCGGTGGA GCAGGGTAC GCCCATGGCG GTGTAGAGTA GCATGCACCT TGAATTACGT CTATAGTGAG GTTACCCAAC CCGATCGCCC CGCGGCGGG1 TGGGCCATCG TTGTTCCAAA TTTCGGCCTA GCTTACAATT Taagagaatt ATTTGGCAAT ပ ပ U SphI 2222 ø CACATTIGIC IGCACTGGGG AIGAAGIGCA CAATTATACA TITAATACGC GATAGAAAC AAAATATAGC GCGCAAACTA GGATAAATTA GGCATGACAG O) GACGAGCGGC TCGTTCAAAC TTTTTATGAT AAAGGAAGAG TGTTTTGCT TCGCCGCATA CCAATTCGCC AAACCCTGGC GCGCATTAAG CGCTCCTTTC GTTCACGTAG TAGTGGACTC ATTTTGCCGA AAATATTAAC TTCTAAATAC TGAGCACTTT TAATTTCTGT GAGGCCCGCA GGGCTCCCTT TACATCGAAC U **>** 回 Ŋ H CHGD CCACTGAAGG ATGAGCTGTA AAGAAGCAGA ATGAGATGGG TTTATAGGG AATTTTAACA TAATATTGAA TTTCCAATGA TCTTACGGAT CACACGACTT ATTTTGGGAT GATTATCATA CGGTGGAGCT CCCTAGCGCC TCTAAATCGG TAGGGTGATG CGTTCTTTAA TTGTTTTT TTTGCCTTCC AGCAACTCGG GTGACTGGGA TAATAGCGAA CCCTGTAGCG ACGAGTGGGT × O ĴΨ × TCACCAGTCA CAGAAAGCA CCTGTTGCCG GTCTTGCGAT TGTAATGCAT GACGTTATTT GCCGGGCLAAG GTTACTAGAT CGATAAGCTT CTAGAGCGGC CCAGCTGGCG ATGGGACGCG CCCGTCAAGC AAAACTTGAT GAACCCCTAT CGAAGAACGT TTACAACGTC CTTGCCAGCG TTGGAGTCCA ATTCTTTGA ATTTAACGCG AATGCTTCAA TTTGCGGCAT AGTTGGGTGC > H e e H Xbar Ħ 2222222 CTGGAGCGTT GATCCTTACA GCGCCCAGAA GCCCAGGGCG GACAGGGTGT GAATAATGCG ATCTCGGTCT AATGTGCGCG AACCCTGATA GCTGAAGATC CCGTATTGAC TGAATGGCGA TCGACCCCAA TTTAACAAAA GTTTTCGCCC CCCCTTTCG GACCGCTACA GCCGGCTTTC CCCTTTGACG TATTCCCTTT CTCATGCCAA CGACGGCATC CCAAACCTGC GGCCGTCGTT đ O) Z ď z Clai τQ ATGAGACAAT AGTAAAAGAT GGTTGAGTAC GGCCGCGTAA AAAGTTTCTT AAGATTGAAT ATAATTAACA CGCGCTCACT TGCAGCACAT TTGCGCAGCC CGCGCAGCGT CGCCACGTTC CGGTTTTTCG ACTCAACCCT AATGAGCTGA TTTTCGGGGA GTGTCGCCCT ATCCTTGAGA CGGTATTATC TTACGGCACC > D P Ø \*\*\*\*\*\*\*\*\*\* O 4 Q GGCCGGTCGC TAAGCATGTA GTATCCGCTC AGAATGACTT TTAATCGCCT TTCCCAACAG CTGGAACAAC CAACATTTCC CGCTGGTGAA CAGCGGTAAG CTATGTGGCG **IGTCATCTAT** CGTATTACG GTGGTGGTTA CTTCCTTTCT ATTTAGTGCT CCCTGATAGA TTGGTTAAAA TAGGTGGCAC Ö **D** O d 1751 1891 2031 2171 841 981 1051 1331 1401 1471 1541 1611 1681 1821 1961 2101 701 771 1261

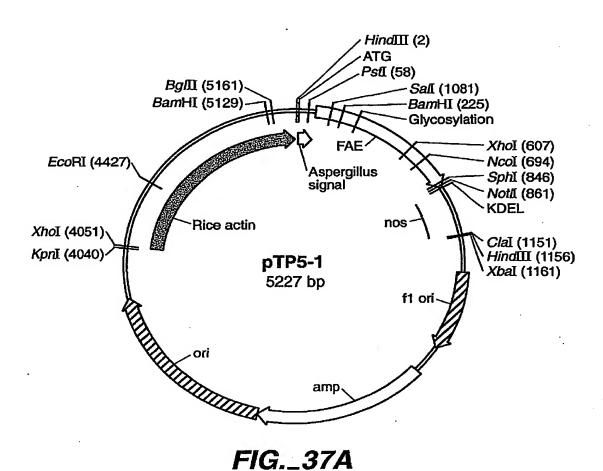
FIG.\_36C

Argeagete eccataacca reagteatha cactegeec aattacter aaggagetaa ccettitit gcacaacta gaggaricate gccetaacac accatacte gaggaricate gccetaacac accategeer trecegeer trecegeer recegeer gettiegee actategacer caggeere credegeer credegeer credegear gattacter gaggager credegear cattacter gaggagere cattacter aacactere cattacter gaggagere recegere accatataa accategare cattacter gaggagere recegere accatataa accategace cattacter gaggagere recegeare gaggagere regaggere credegeare credegeare credegeare credegeare gaggere aacatacter gaggagere regaggere regaggere gaggagere regaggere gaggagere aacatactaac gaggagere aacatactaac gaggagere aacatactaac gaggagere aacatactaac gaggagere credegere aacatactaac gaggagere credegere aacatactaac gaggagere credegere gaggagere regaggere gaggagere ga	GGGTATTTT	TTGCAGCGAT	CTTGCCCCCG	3 ATTAAAATAG C	CAATAATAAG	CCACAATGAA	CAGGCGAATT	4481 1
Argchatget GCCataracca reagreatha cactaccec arctractic reacaacean Aaggactrac GCCGAACTAC GACGAGCOTG ACACCACGAT GATTAATAC TGAATTAACT GGCGAACTAC GACGAGCOTG ACACCACGAT GCCTGAGCA ATGGCAACAA GTTATTAACT GGCGAACTAC GACGAGCOTG CTCCGGGCA CAATTAATAC AGCGTGGGC CTCGGGCCCTT CCGGCTGCT GATTAATACG ACATGATACT AGCGTGGGC CCCAACGGCAAC TCCGGGCAA TGGCAACAA TTAAAAGGAT CTGGGTAACA ACACAGAAACT TCTTAAAAGGAT TTAAAACTT TTAAAAGGAT CTAGGTGAAA ACCAAGTTTAC TCTTAAAATTAC TCTTAAAACTT TTATTCCGAA AACACAGAAA ACCAAGTTACC GCCTAAAATC CTTTAAAACTT GCTGCCTGC TCAAAAAAAA ACCACCCCTA CGCTGATTAC CACCAAAATC CTTTAAAACTT TTTTTCCGAA GGTAACAAAAAA ACCACCACAAAAC CAAATACTTTCT AGGCCACCAC TCAAAAAAAA ACCACCACAAA GGCAAAAAC CTTTAAAAACTT GCTGCCCACCAC GAACACAAAAAA ACCACCACAAAAC CAAAAACTT GCTGCCCACCAC GAACACAAAAAA ACCACCACACA GGCACAAAAC CTTTAAAAACTT GCTGCCCACCAC GAACACAAAAAA ACCACCACACA GGCACAAAAC CTTTAAAAACTT GCTGCCACCAC GAACACAAAAAA ACCACCACACA GGCACAAAAC CTTTAAAAACTT GCTGCCCACCAC GAACACAAAAAAA ACCACACACACA GGCACAAAAC CTTTAAAAACTT GCTGCCCACCAC GAACACAGAAA ACCACACAAAA GCCCACACACA	(さくごうかみ かけいかか	TTGCAGCGAF	STATE COCCO	ATTAAAATAG	СААТВАТААС АТТААААТАС	CAGGGAAIT CCACAAIGAA	CAGGCGAATT	4481
Argcagarger GCCATAACCA TGAGTGALAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CGCCTTTTTT GCACAACGTG GGGGATCAIG TAACTCGCCT TGATCGTTGG AAGGAGCTAA CGCCTTTTTTT GCACAACGTG GGGGATCAIG TAACTCGCCT TGATCGTTGG CTACTAAAGGAT GACGAGCGTG ACCGGCCAA GGGTAAACGT GATTAAAACT AGCGTGGGTC TCCCGGCCCT TCCCGGCCAA TGGTAAGCC TCCCGGTATCG CACGACGGG AACTAAAAAAAA CTATGCAGTAC TCATGAATTAC CTTTAAAACGTG TTAAAAGGAT CTAGACCCC TAGAAAAAAA CCACCGCTACATCAT TTAAAAGGAT CTAGACCCC TAGAAAAAAAA ACCACCGCTACATCAC TTAAAAGGAT CTAGACCCC TAGAAAAAAAA ACCACCGCTACATCAC TTAAAAGGAT CTAGACCCC TAGAAAAAAAA ACCACCGCTACATAC TTAAAAGGAT CTAGACCCC TAGAAAAAAAA ACCACCGCTACATAC TTAAAAGGAT CTAGACCCC TAGAAAAAAAAAAAAAAAA	AATATATATT	TTTTGAGAAA	TIGACATAAT	ATATGCTAAT	TAAAAAACCC	GAAATATCTT	ATTTGTATAA ECORI	4411
Argcagter GCCTTTTT GCAACATG GGGGATCATG TAACTCGCT TGATCGTGG AAGGAGTTAACT GCAACGAT GCCTGTAGG TGATCGTGG TGATCGTGG TGATCGTGG TGATCGTAG TGATCGTGG TGATCGTG TGATCGTG TGATCGTG TGATCGTG TTTAAAGGTTG TTTAAAAGGT TTTATCGAA AAGCTTGT TGAGTGAAAAAAAACTG TTTAAAAAGGT TTTAAAAAGGT TTTATCCGAA AACCACTGATA GGGTTGGT GGGTTGGT GGGTGGT TTTATCGAAAAAAAAAA	ATACAGAGGG	GCTTTTGTAA	TAAGTTCGTT	AGTCGGTTTT	TCTGTATTTG	GAAATGCATA	TCGCGATTTG	4341
ARGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGCC AACTTACTTC TGACAACGAA AAGGAGGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTGGCA TGATCATGGA ATGGCAACAA GACCACGAA CACTTAAAACA CATTAAAACT GGCGAACTAC GACGACGCT GATTAAAACAA ACTATAAAACT GGCGAACTAC CTGACCAGA CACTTAAAACAA CATTAAAACT GGCGAACTAC CTGACCAGA TGGTAAACC TAACACACAAATC CTGCTTACCAAACT TAAAAACTTTTTTTAAAAACTTTTAAAAACTTTTTAAAAACTTTTTAAAAACTTTTTT	TTAAGTTTAT	GAATTGGTTT	ATTTTTGTAT	TTGATACGTC	TGTCGGTACT	GAGGATGTTT	TATAAAATT	4271
Argcagatget gecataacca tragtaataa cattecege aacttactee tragtecaa aaggagettaa cestitaataa gacaacata gagaatcatg aacteaacaa aaggaactaa gacaacataa gagaatcatga acteaataaa actaaataa actaaataa actaataaa actaaataa actaaaataa attaaaataa actaaaataa actaaaataa actaaaataa actaaaataa actaaaaaa actaaaaaaa actaaaaaaa actaaaaaaa actaaaaaa actaaaaaaa actaaaaaaa actaaaaaaaa	TTTCTACTAT	AATTTACTCT	CCCAAAGTGA	TAAAAGGTGG	ATATCGGTAA	TATAAGTAAA	TAAAAGGTGG	4201
ARGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACAC GATTAATAG ACTGCAACAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA CANTTAATAG ACTGCATGG ACTGCAGGAC CACTTCTGGC TTCGGGCCTG GGTTTAATAGG CACTGAACTT AGCGTGGGTC TCCCGGTAAC ACTGCACACAC ACATGACTTC AGCGTGGGTC TCCCGGTAAC TTCGGGCTGA ACATGACTC GATTAAACTT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GATTAAACTT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAATAC TTTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCACAGTG TCCTGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAATAC TTTAAAAGGAT CTGTAGCACAC CCAGCGGTG TTTAAAACTT TTAAAAGGAT CTGTAGCACAC GCCCACACTAC GACCACAATAC TTTAAAAGGAT CTGTAGCACAC GCCCACACTAC GACCACAAACT TTTAAAAGGAT CTGTAGCACAC GCCCACACTAC GACCACAAAC TTTAAAAGGAT CTGTAGCACAC GCCCACACTAC GACCACAAAC TTTAAAAGGAT CTGTAGCACAC GCCCACACTAC GACCACAAAC TTTAAAAGGAT CTGTAGCACAC AGCCCAGCTA GACCACAAAC TCCTACAGCGT GACTAAAAAGA CCACCACACAC GACCCACACAC GCCCACAGTA GCCCACACAC GCCACACACAC GCCCACACAC GCCCACACAC GCCCACACAC GCCCACACAC GCCCACACAC GCCCACACAC GCCCACACAC ACACACAC	AAACATCAGT	TCAAAAGTGA	GATTACCTGG	ACAAAGGTAA	CCAAAATAAA	TCGGGATAGT	GAGAAGAGAG	4131
ARGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACAT GCCAGACTAC TGAATGAACG CACTGCGCTG ACACCACAT GCCAGACTAC ACTATTAACT GGCGACTAC TTACTCTAGC TCCCGGCA GATTAATAG ACTGCATGG ACGTGGGGC CTCGGCCCT CGGGCCCAG TGATTAATAG ACTCGATGC CACGACGGG AGTCAGGCAC TCTGGCCCTT CGGGCCCAG TGATTAATAGT TAAAAGGAT CACGATGCT CACGATGCT GATTAAAAGGAT CACGATGCT CACGATGCT TAAAAGGAT CACGATGCT CACGATGCT TCCCGGTAATACT TTAAAAGGAT CACGAAAAAAAA ACCACCGGTAA CCACGGGGG TTCAGCAGAAAAAAAAAA	TCATATGCTT	TCGAGGTCAT	2222222555	GCTGGGTACC	GGGAACAAAA	CCTCACTAAA	CGCAATTAAC	4061
Argcagicta cccitation teachana cachegece aactiactic teachacdar aaggagetaa cccitititi gcachacate gegeaitatic transfering accepted acacegory gegeafect cataccaac argecacca catteratas actegoracaa actatiaact gegeafect cataccaaca catteratas actegoraca actatiaact gegeafect catteratac catteratas actatiaact gegeafect catterates catteratas actatiaact accepted actegory coectegory gettaratac teathacter catterates actatiaact transfering actatiaact transfer catteratas actatiaact transfer catteratas actaticate actatiaact transfer catteratas actaticate actatiaact transfer catteratas actaticate actatiaact transfer catteratas actatiaact transfer catteratas actatiaact catteratas actatiaact catteratas actatiaact catteratas catteratas actatiaact catteratas actatiaact catteratas actatiaact catteratas actatiaact catteratas actatiaact catteratas actatiaact catteratas actaterate actaticatas actatiaact catteratas actaterate a		*****		******				
Argcagtect gccataacca reagteataa cacteceecc nacttactec reacaacean aageaagetaa cogcittitit gcacaacate gegearcate raattacte gagaactaca argcaacaa actaataaca actacaacaa actaataaca cataataaca actacaega cataataaca actacaega cataataaca actacaega cataataaca actacaega cataataaca actacaega cataataga cataataaca aageataga actaataga cataataaca atacaacaga cataataga actaataga actaataga cataataga attaataga cataataga attaataga actaataga actaataga actaataga actaataga actaataga actaataga actaataga actaataga actaataga attaataga at		thoI	~	KpnI				
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGC AACTTACTTC TGACAACGAT AAGGAGCTAA CGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAACACA CGACACGATG GGGGATCATG TAACTCGCCT TGATCGATGG TGAATGAACACACACACACACACACACACACACACACACA	ACGCCAAGCG	GACCATGATT	AAACAGCTAT	TTCACACAGG	GGATAACAAT	AATTGTGAGC	TGTTGTGTGG	3991
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACCCACGAA CAATTAATAG ACTGGATGG AGGGAACACA CAATTAATAGC TAACTCGGCAC TGAGGCACAAAAAAAACTT CCGGCTGGCAC CACTTGCAGGCA CAATTAATAGCA CACTTCTGG CTCGGCCCTT CCGGCTGGCA CAATTAATAGCA CTCGGTATCG CTCGGCCCTT CCGGCTGGCA CAATTAATAGCA CTTAAGGCAC CCAAGTTTTTC TAAAAGCAT CCACTGAACGTG AGCCATGAACGTG ACCCGGTAGA CCAAATTAAAACTT CTTAAAAGCAT CTTAAAAACTT TTAAAAGCAT CTTAAAAACTT CTTAAAAACTT TTAAAAGCAT CTTAAAAACTT CTTAAAAACTT CTTAAAAACTT CTTAAAAAACT CTTAAAAAACT CTTAAAAAACT CTTAAAAAACT CTTAAAAAACT CTTAAAAAACT CTTAAACACA GTTAACCAGAATAC CTTAACACAAAAAAAAAA	CCGGCTCGTA	CITTATGCTT	AGGCTTTACA	TAGGCACCCC	GCTCACTCAT	ATGTGAGTTA	AACGCAATTA	3921
ANGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGC AACTTACTTC TGAGAGGAT AAGGAGCTAA CGCTTTTTT GCCAACATG GGGGATCATG TAACTCGCCT TGATGGTTGG TGAATGAAGC CATACTACT GCCGACGGT ACACCACGAT GCCTGTAGCA ATGCAACACA ACACTACTAG GGGGATCATG ACACTACTAG ACACTACTAG ACACTACTAG ACACTACTAG ACACTACTAG ACACTACTAG ACACTAGATGA ACGAATAGA ACCTTAATAGGA ATTGCAGGA CTATGGAGGAC TGGGGCCAGA TGGTAAGGTC TCTGGGTGAAGT CTTTAAAAGGAT CTATGGATGA ACGAATAGA ACCTTAATAGAACTT TCTGAGGA AACCGGTAAA ACCTTTTTTTTC TAAGATTGC TCATGTAAAAGA ACCTTTTTTTTC ATAAAAGGAT TCTTGAGATC TTTAAAAAGGAT TCTTGAGAACT TTTAAAAGGAT TCTTGAGAACT TTTAAAAAGGAT TCTTGAGAACT TTTAAAAAGGAT TCTTGAGAACT TTTAAAAAGGAT TTTAAAAAGGAT TTTAAAAAGGAT TTTAAAAAGGAT TTTAAAAAGGAT TCTTGAGAACT CTTTAAAAAGAT GCTTAATACT TCTTGAGAACT CTTTAAAAAGAT TCTTGAGAACT CTTTAAAAAGAT TTTTTCTGAA ACCTGACACAC GCTTACCAATAC GCTTAACTATATAC TTTAAAAAGAT GTTAAAAAGAT TTTTTTTTTT	CAGTGAGCGC	GGAAAGCGGG	TTTCCCGACT	GCACGACAGG	AATGCAGCTG	CCGATTCATT	CGCGCGTTGG	3851
Argeagectaa CCecttaacca Teagearaa Cactegegec Aacttacter Teacaacgar Aaggaagettaa CCecttater Gaccaacac Gaccaacac Gaccaccac Caccaccac Caccaccac Caccaccac Caccaccac Caccaccac Gaccaccac Caccaccac Gaccaccac Gaccacacac Gaccacacac Gaccacac Gaccacacac Gac	CGCCTCTCCC	ATACGCAAAC	AGAGCGCCCA	AGGAAGCGGA	TCAGTGAGCG	GCGCAGCGAG	GAACGACCGA	3781
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACAG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC ACCCACGAT GCCTGTAGCA ATGCCACAA ACTACTAAACGA CACTTAAAACGA TGCGAACTAC TTACTCTTAGC TTCCCGGCAA TGGTAAGCC TCCCGGTACAACTAC TTACTCTTACA CCGCTGCTAACTAC TTACTCTTACA CCGCTGCTAACTAC TCGGCTGCAACTAC TCGGCTGCAACTAC TCGGCTGCAACTAC TCGGCTGCAACTAC TCGGCTGCAACTAC TCGGCTGCAACTAC TCGCGCTGAACTAC TCGCGGCTAACCT TCGCGCTGAACTAC TCGCGCTGAACTAC TCGCGCTGAACTAC TCGCGGCTGAACTAC TCGCGGCTGAACTAC TCACGGTGAACTAC TCACGGGGGGT TCGTGCACCACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGGGGGGTTTAACTAC TCACGGGGGGTTTAACTAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCAC TCACGCTACACCACACACACACACACACACACACACACAC	CGCCCAGCC	TGATACCGCT	TTGAGTGAGC	ATTACCGCCT	GGATAACCGT	CTGATTCTGT	GCGTTATCCC	3711
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAACTCG CATACCAACA GACGAGCGA CAATTAATAG ACTGGATGGA GACTGAACTAC CATACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GATTTATTTCC AGCGTGGCA CAATTAATAGC CACTGGATGA CAATTAATAGC CACTGGATGA CAATTAATAGC CACTGGATGA CAATTAATAGC TCCGGTGACTA CAATTGCAGCAC TGGGGCCCAGA TGGTAAGCTC TCCGGTGACA CAATTAATAGC CACTGGATGA AAGCGTGAACTC TCAAGTTTTTC ATAAAAGGAT CAATTAATAC CAAGTTTTTCCGAA AACTGTGAAAAC CAAGCGGCGCTA CAAAGAATC CTTTAAAAAGCAT CAAAGAATC CAAGCGGCGCTA CAAAGAATC CAAGCGGCGCTA CAAAGAATC CTTTAAAAACTC AAACAAAAAA ACCACCGCTA CAAAAGAATC CTTTAAAAACTC CAAGTTTTTTTTTT	GTTCTTTCCT	TTGCTCACAT	TGCTGGCCTT	CCTGGCCTTT	TTTTACGGTT	AACGCGGCCT	AAACGCCAGC	3641
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAC GACGAGCGTG ACCCACGAT GCCTGTAGCA ATGCCAACAAC GACGAGCGTG TTCCCGGCAA CAATTAATAAG ACTGGATCGA CATTAATAAGC TCGGGTACC TCCGGTATCT TGATGGATCA CATTAATAAGCAC TCGGGTACC TCGGGTACC TCCGGTATCT TAAAAGGATC TCGGGTAACTC TCGGGTAACTC TCGGTAACTC TCGGTAACTC TCGGTAACTC TCGGTAACTC TCGGTAACTC TCATAGATCA TTAAAAGGAT TTAAAAGGAT CTAGGTGAAG ATCACTTTTC AGATCGTGAAG TCCTTTTTC AACCATTGC GATCAAAATC TTAAAAGGAT TCTTGAGTGA ACCACTGAAAAAA ACCACCGCTA CAACGAAAAAA ACCACCGCTA CCAGCGGTG TTTTTTCCGAA ACCACAAAAAA ACCACCGCTA CCAGCTGAC CTTTTTTTCCGAA GGTAAACTC TCAGTAGCAC GGTTGGAACTC TCAGTAGCAC GGTTGGAACTC GTTTTTTTCCGAA GGTCAAAACTC GTTTAACACACC TCGTAACTCC GGTTGGACTC CTTTTTTTTTT	GCCTATGGAA	GGGGGCGGA	ATGCTCGTCA	GATTTTTGTG	CTTGAGCGTC	CCACCTCTGA	TCGGGTTTCG	3571
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAC GACGAGCGTG ACCCACGAT GCCTGTAGCA ATGCCAACAA GACTATTAAACT GGCGAACTAC TTACTCTAGC TCCGGCCAA CAATTAATAG ACTGGATGGA GATTAATAGCAC TCCGGTATCG TGGCTGGCAC TCCGGTATCG TGATTAATAGGAT TCCGGTATCG TGATTAATAGCAC TCCGGTATCG TGGCTGGCAC TCCGGTATCG TGATTAATAGGAT TTAAAAGGAT CTAGGTGAAG TCATTGTTTTCC TTAAAAGGAT CTAGGTGAAAACTT TCCTAACTGC TCAGAACTGC TAAAAGCATCG TCAGAACTGC TAAAACTGC TCAGAAAACTT TCCTAAAAGGAT TCTTTCCGAA ACCACCGCTA CCAGCGGTGG TTTTTTCCGAA GGTAACTGC TCAGAACTGC TCAGAACTGC TCAGAACTGC TCAGAACTGC TCAGAACTGC TCAGCAGATCC CTTTTTTTCCGAA GGTAACTGC TCAGAACTCC TCAGAACTCC TCAGAACTGC TCAGCAGATCC CTTTTTTTCCGAA GGCTAAAACTC GCTTCAGAACTC CTTTTTTTTTT	TATAGTCCTG	CTGGTATCTT	GGGGAAACGC	GAGCTTCCAG	GCGCACGAGG	GAACAGGAGA	GGCAGGGTCG	3501
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAC GACGAGCGTG ACCCACGAT GCCTGTAGCA ACTACTACT GACGACCTAC CTCCCGGCAA CAATTAATAG ACTGGATGGA GTTGGATCAC TGACGAGCAC CACTTCTGGG AGTCGTAGC CTCGGCCTGC CTCGGCCTGC CTCGGCTACC TGCGGTATCG TGGCTAAGCCC TCCCGTATCG AGGCTGCGGG AGTCAGGGCAA CTATGCAGGCA TGGTAAGCCC TCCCGTATCG AGGCTTGGA ACCGATGGCAA CCAAATAGCTG TTAAAAGGAT CTAGGTGAAG CCACTTTTTC AACTGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAATC CTTTAAAAGGAT CTAGGTGAAAAGT CAAAGGATCT TCTTGAGATC CTTTTTTCCGAA GGTAACTGG TTCAGCAGAG CGCAGATACC CTTTTTTTCT GATTTTTCCGAA GGTAACTGGC TTCAGCAGCTC CTTTTTTTCT GATTTTTCCGAA GGTAACTGG TTCAGCAGCTC CAGCGGGGGT TTGTTTTTCT GGTAACTGC TAATCCTGTT GGCCCACCAC TAATCCTGTC TAATCCTGTT GGCCCACCAC TAATCCTGTT GGCCCACCACCAC TCTTTTTTTTTT	TCCCCTAACC	CGGACAGGTA	GGGAGAAAGG	GCTTCCCGAA	AAAGCGCCAC	GAGCTATGAG	CCTACAGCGT	3431
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGCCAACA GACGAGCGTG TTACTCTAGC TTCCCGGCAA CATTAATAG ACTGGATGGA GTTGCAGGAC CACTTCTGG TGATTAATAG TGATAAATCT GGCGTGGCC TCCCGGTATC TGGGCCAGA TGGTAAGCC TCCCGTATCG AGGCACTGG AGGCGTGC AGGCGTGC AGGCGTGC TCCCGTATCG TGGCGTGGA ACCGATGGCAA CCAAATAGA ACCGAAATAC TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAATGC CTTTAAAAGGAT CTAGGTGAAAAAGAT CAAAGGATCT TCTTGAGATC CTTTATTTCCGAA ACCACCGCTA CCAGGGGTGG TTTTTCCGAA ACCACAAAAG CGCAGATACC CTTTTTTCCGAA ACCACAAAACTGC TCAGCAGATCC CTTTTTTTCCGAA GGTAACTGGC TTCAGCAGATC CTCTTAGATTCT GATTTTTCCGAA GGTAACTGGC TTCAGCAGATC CTCAGGATCT CTTTTTTCCGAA GGTAACTGGC TTCAGCAGATC CTCAGCAGTT AGGCCCCACCACCACTA CTCAGCAGTT CTTTTTTCCGAA GGTAACTGCC GGTTTAACCGAATAC CTCAGCAGTT CTTTTTTTTTT	AACTGAGATA	ACCTACACCG	GGAGCGAACG	AGCCCAGCTT	TCGTGCACAC	AACGGGGGGT	GGTCGGGCTG	3361
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ACTGCAACAC GACGAGCGTG CTCCCGGCAA CAATTAATAG ACTGGATGGA GTTGCAGGAC CACTTTTTGC GCCTGCCT CGGCTGGCT CGGCTGCTC CGGCTGGCT	AAGGCGCAGC	GTTACCGGAT	CAAGACGATA	GGGTTGGACT	GTGTCTTACC	GCGATAAGTC	GCTGCCAGTG	3291
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTTTTTT	ACCAGTGGCT	TAATCCTGTT	CICGCICIGC	GCCTACATAC	CTGTAGCACC	TTCAAGAACT	AGGCCACCAC	3221
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAAGCTAA CCGCTTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACCCACGAT GCCTGTAGCA AAGGAACAAAAAAAAAA	AGCCGTAGTT	CTTCTAGTGT	AAATACTGTC	CGCAGATACC	TTCAGCAGAG	GGTAACTGGC	TTTTCCGAA	3151
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACCCACGAT GCCTGTAGCA ATGCCAACAA GACGAACTAA GACGAACTAC TACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATGAATCG AGCGTATCG AAGCGTATCG AAGCGTGAA CCAAATTATAAAAGGTGAAGAACTTAAAAAGGATC TTAAAAAGGATC CTTAAAAAGGATC CTTTAAAAAGCTG ACCTGAAAATC CTTTAAAAAGCTG CTTTAAAAAGCTG CCACTGAAAATC CTTTAAAAACCTG CTTTAAAAAGGATC CCACTGAAAATC CTTTAAAAACCTG CACACTGAAAATC CTTTAAAAACCTG CTTTAAAAAGTTTTTTT CTAAAAGGATCT TCTTGAGATC CTTTAATTTTTTTTTT	CTACCAACTC	GGATCAAGAG	TTTGTTTGCC	CCAGCGGTGG	ACCACCGCTA	AAACAAAAA	TGCTGCTTGC	3081
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGGGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCTTG TGAATGAAGC CATACCAAAC GACGAGCGTG ACCCACGAT GCCTGTAACA ATGGCAACAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTAATGC TGATAAAATCT AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGCCCAGA TGGTAAGCCC TCCCGTAATCG CACGACGGG AGTCAGGCAA CTGTAGGATGA ACGAAATGA CAGATCGCTG AGATAAGGTGC AAGCATTGC TCATAAAAACGTG AACTGTCAGA TCATAAAACTT GATAAAACCTG TTAAAAAGCAT CTTAAAAACCTG CCTTAAACCTG	GCGCGTAATC	CTTTTTTCT	TCTTGAGATC	CAAAGGATCT	TAGAAAAGAT	TCAGACCCCG	CCACTGAGCG	3011
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA GATGCATGGA CAATTAATAG ACTGGATGGA GATGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTAATTGC TGATAAATCT AGCGTGGGTC TCGCGGTATC ATGCAGGCAC TGGGGCCCAGA TGGTAAGCTC TCCCGGTATCG TGGGGCCCTT CCGCTAATGAATCG TCGCGGTATCG AAGCGTATCG TCGCAAGTTTAC TCATAATAAAACTT AAGCATTGGT AACCATCAGA CCAAGTTTAC TCATAATAAAACTT TCATAAAAACTT TCATAAAAACTT TCATAATAAAACTT TCATAAAAACTT TCATAAAAACTT TCATAAAAACTT TCATAATAAAACTT TCATAAAAACTT TCATAAAAACTT TCATAATAAAACTT TCATAAAAACTT TCATAAAAAACTT TCATAAAAAACTT TCATAAAAAACTT TCATAAAAAACTT TCATAAAAAACTT TCATAAAAAACTT TCATAAAAAACTT TCATAAAAAAACTT TCATAAAAAAAA	ない。これはいいいのでは、	CCTTAACGTG	GACCABAATC	ATAATCTCAT	ATCCTTTTG	CTAGGTGAAG	TTAAAAGGAT	2941
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTAATTGC TGATAAATCT AGCGTGGGTC TCGCGGTATC ATGCAGGCAC TGGGGCCCAGA TGGTAAGCCC TCCCGTAATCG CACCAGAGCAA CCACACGCTG AGATAGCTG AGATAGCTG	CA thribdahaba A an	TTTAAAACTT	TTTAGATTGA	TCATATAC	CCAAGTTTAC	AACTGTCAGA	AAGCATTGGT	2871
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA ACTGGATGGA GATTAAATA GCGGAACTAC TACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTAATTGC TGATAAATCT AGCGTGGGTC TCGCGGTAACCC TCCCGGTATCG	CTCACTGATT	AGATAGGTGC	CAGATCGCTG	ACGAAATAGA	CTATGGATGA	AGTCAGGCAA	CACGACGGGG	2801
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT GGTTTATTGC TGATAAATCT	TAGTTATCTA	TCCCGTATCG	TGGTAAGCCC	TGGGGCCAGA	ATTGCAGCAC	TCGCGGTATC	AGCGTGGGTC	2731
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA	GGAGCCGGTG	TGATAAATCT	GGTTTATTGC	CCGGCTGGCT	CTCGGCCCTT	CACTTCTGCG	GTTGCAGGAC	2661
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTT GCACAACATG GGGGATCATG TAACTGGCCT TGATCGTTGG TGAATGAAGC CATACCAAAC GACGAGGGG ACACCACGAT GCCTGTAGCA ATGGCAACAA	GGCGGATAAA	ACTGGATGGA	CAATTAATAG	TTCCCGGCAA	TTACTCTAGC	GGCGAACTAC	ACTATTAACT	2591
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT AAGGAGCTAA CCGCTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG	CGTTGCGCAA	ATGGCAACAA	GCCTGTAGCA	ACACCACGAT	GACGAGCGTG	CATACCAAAC	TGAATGAAGC	2521
ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT	GAACCGGAGC	TGATCGTTGG	TAACTCGCCT	GGGGATCATG	GCACAACATG	_	AAGGAGCTAA	2451
	ยบบลยยลยยบ	中のなるなのなの中	AACTTACTTC	CACTGCGGCC	TGAGTGATAA	_	ATGCAGTGCT	2381

## FIG.\_36E

TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG AATTTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC

5181 5251 5321



SUBSTITUTE SHEET (RULE 26)

AAGCTTAACA TGAAGCAGTT CTCCGCCAAA CACGTCCTCG CAGTTGTGGT GACTGCAGGG CACGCCTTAG 四 Ø PetI 田 Ħ E A ď Ø Œ × × Hindrin

4

7

CAGCCTCTAC GCAAGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC

D L C N I P S T I I K G E K I Y N S Q T GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAACT BEMHI · A Y A TGCCTACGCC 141

gacattaacg gatggatcct ccgcgacgac agcagcaaag aaataatcac cgtcttccgt ggcactggta Ö 辉 [24 Н 闰 ĸ Ω Ø × A Z K A H M Н A

GTGATACGAA TCTACAACTC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA 2 U H H Þ A Ba ø ⊳ ρι × × 되 þ H O H ø ס E C บ • 211 281

CAGCAGGTTA GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG 4 D H O > 0

CACTCACTGC CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG Ø p; H 351 421 491 • G N Q A F A S Y M N D A F Q A S S P D T T Q Y CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT NGOI

TTCCGGGTCA CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG Ħ **7** () O A V E E E E Ø H (J A Æ H H F1 X2 < 631

CTGGAGCGTT GATCCTTACA GCGCCCAGAA CACATTTGTC TGCACTGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC © ₽4 ≥4 e H N > D O פ O GTGTAGAGTA ບ 701 771

		TCGTTCAAAC	TAATTTCTGT	TTTTATGAT	GGATAAATTA			CCAATTCGCC	AAACCCTGGC	GAGGCCCGCA	GCGCATTAAG	CGCTCCTTTC	GGGCTCCCTT	GTTCACGTAG	TAGTGGACTC	ATTTTGCCGA	AAATATTAAC	TTCTAAATAC	AAAGGAAGAG	TGITTIECT	TACATCGAAC	TGAGCACTTT	TCGCCGCATA	GGCATGACAG	TGACAACGAT	TGATCGTTGG	ATGGCAACAA	ACTEGATEGA	TGATAAATCT	TCCCGTATCG
		AAGAAGCAGA	GATTATCATA	ATGAGATGGG	GCGCAAACTA			CGGTGGAGCT	GTGACTGGGA	TAATAGCGAA	CCCTGTAGCG	CCCTAGCGCC	TCTAAATCGG	TAGGGTGATG	CGTTCTTTAA	TTTATAAGGG	AATTTTAACA	TTGTTTTTT	TAATATTGAA	TTTGCCTTCC	ACGAGTGGGT	TTTCCAATGA	AGCAACTCGG	TCTTACGGAT	AACTTACTTC	TAACTCGCCT	GCCTGTAGCA	CAATTAATAG	GGTTTATTGC	TGGTAAGCCC
	* 1 M Q	CCACTGAAGG ATGAGCTGTA	GTCTTGCGAT	GACGTTALT	AAAATATAGC :	XbaI	22222	CTAGAGCGGC	TTACAACGTC	CCAGCTGGCG	ATGGGACGCG	CTTGCCAGCG	CCCGTCAAGC	AAAACTTGAT	TIGGAGICCA	ATTCTTTGA	ATTTAACGCG	GAACCCCTAT	AATGCTTCAA	TTTGCGGCAT	AGTTGGGTGC	CGAAGAACGT	GCCGGGCAAG	CAGAAAAGCA	CACTGCGGCC	GGGGATCATG	ACACCACGAT	TTCCCGGCAA	CCGGCTGGCT	TGGGGCCAGA
	PLKI	CCACTGAAGG	CCTGTTGCCG GTCTTGCGAT	TGTAATGCAT GACGTTATTT	GATAGAAAC AAAATATAGC Hindili	Clar	222222	CGATAAGCTT	GGCCGTCGTT	CCCCTTTCG	TGAATGGCGA	GACCGCTACA	GCCGGCTTTC	TCGACCCCAA	CCCTTTGACG	ATCTCGGTCT	TTTAACAAAA	AATGTGCGCG	AACCCTGATA	TALTCCCLTT	GCTGAAGATC	GTTTTCGCCC	CCGTATTGAC	TCACCAGTCA	TGAGTGATAA	GCACAACATG	GACGAGCGTG	TTACTCTAGC	CICGGCCCIT	ATTGCAGCAC
Noti	A A E			ATAATTAACA	TTTAATACGC	ដ	1 1 1	GTTACTAGAT	CGCGCTCACT	TGCAGCACAT	TTGCGCAGCC	CGCGCAGCGT	CGCCACGTTC	TTACGGCACC	CGGTTTTCG	ACTCAACCCT	AATGAGCTGA	TTTTCGGGGA	ATGAGACAAT	GTGTCGCCCT			CGGTATTATC	GGTTGAGTAC	GCCATAACCA	CCCCTTTTTT	CATACCAAAC	GGCGAACTAC	CACTTCTGCG	TCGCGGTATC
2 1	<b>₹</b> > a			TAAGCATGTA	CAATTATACA			TGTCATCTAT	TCGTATTACG	TTAATCGCCT	TTCCCAACAG	GTGGTGGTTA	CTTCCTTTCT	ATTTAGTGCT	CCCTGATAGA	CTGGAACAAC	TTGGTTAAAA	TAGGTGGCAC	GTATCCGCTC	CAACATITCC	CGCTGGTGAA	CAGCGGTAAG	CTATGTGGCG	AGAATGACTT	ATGCAGTGCT	AAGGAGCTAA	TGAATGAAGC	ACTATTAACT	GTTGCAGGAC	AGCGTGGGTC
Sphi	A C T W	GCATGCACCT	ATTTGGCAAT	TGAATTACGT	TAGAGICCCG			TCGCGCGCGG	CTATAGTGAG	GTTACCCAAC	CCGATCGCCC	CGCGGCGGGT	GCTTTCTTCC	TAGGGTTCCG	TGGGCCATCG	TTGTTCCAAA	TITCGGCCIA	GCTTACAATT	ATTCAAATAT	TATGAGTATT	CACCCAGAAA	TGGATCTCAA	TAAAGTTCTG	CACTATTCTC	Taagagaatt	CGGAGGACCG	GAACCGGAGC	CGTTGCGCAA	GGCGGATAAA	GGAGCCGGTG
		841	911	186	1051			1121	1191	1261	1331	1401	1471	1541	1611	1681	1751	1821	1891	1961	2031	2101	2171	2241	2311	2381	2451	2521	2591	2661

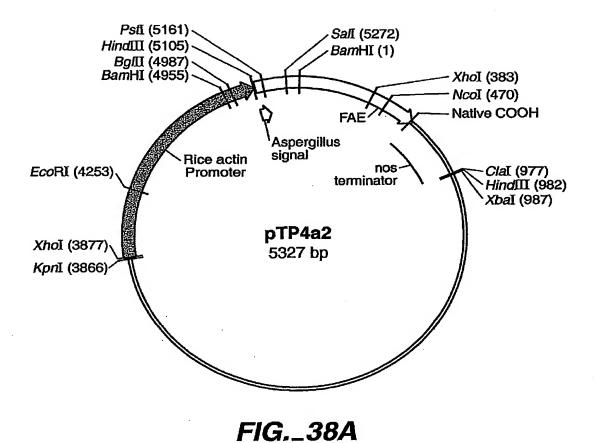
## F/G.\_37C

2731	TAGTTATCTA	CACGACGGGG	AGTCAGGCAA	CTATGGATGA	ACGAAATAGA	CAGATCGCTG	AGATAGGTGC
1087	CTCACTGATT	AAGCATTGGT	AACTGTCAGA	CCAAGTTTAC	TCALALATAC	TTTAGATTGA	TTTAAAACTT
2871	CATTTTTAAT	TTAAAAGGAT	CTAGGTGAAG	ATCCTTTTTG	ATAATCTCAT	GACCAAAATC	CCTTAACGTG
2941	AGTTTTCGTT	CCACTGAGCG	TCAGACCCCG	TAGAAAAGAT	CAAAGGATCT	TCTTGAGATC	CITITITICI
3011	GCGCGTAATC	TGCTGCTTGC	AAACAAAAA	ACCACCGCTA	CCAGCGGTGG	TTTGTTTGCC	GGATCAAGAG
3081	CTACCAACTC	TTTTCCGAA	GGTAACTGGC	TTCAGCAGAG	CGCAGATACC	AAATACTGTC	CTTCTAGTGT
3151	AGCCGTAGTT	AGGCCACCAC	TTCAAGAACT	CTGTAGCACC	GCCTACATAC	CICGCICIGC	TAATCCTGTT
3221	ACCAGTGGCT	GCTGCCAGTG	GCGATAAGTC	GTGTCTTACC	GGGTTGGACT	CAAGACGATA	GTTACCGGAT
3291	AAGGCGCAGC	GGTCGGGCTG	AACGGGGGGT	TCGTGCACAC	AGCCCAGCTT	GGAGCGAACG	ACCTACACCG
3361	AACTGAGATA	CCTACAGCGT	GAGCTATGAG	AAAGCGCCAC	GCTTCCCGAA	GGGAGAAAGG	CGGACAGGTA
3431	TCCGGTAAGC	GGCAGGGTCG	GAACAGGAGA	GCGCACGAGG	GAGCTTCCAG	GGGGAAACGC	CIGGIATCIT
3501	TATAGICCIG	TCGGGTTTCG	CCACCTCTGA	CTTGAGCGTC	GATTTTTGTG	ATGCTCGTCA	GGGGGCGGA
3571	GCCTATGGAA	AAACGCCAGC	AACGCGGCCT	TTTTACGGTT	CCTGGCCTTT	TGCTGGCCTT	TTGCTCACAT
3641	GTTCTTTCCT	GCGTTATCCC	CTGATTCTGT	GGATAACCGT	ATTACCGCCT	TTGAGTGAGC	TGATACCGCT
3711	CGCCGCAGCC	GAACGACCGA	GCGCAGCGAG	TCAGTGAGCG	AGGAAGCGGA	AGAGCGCCCA	ATACGCAAAC
3781	CGCCICICCC	CGCGCGTIGG	CCGATTCATT	AATGCAGCTG	GCACGACAGG	TTTCCCGACT	GGAAAGCGGG
3851	CAGTGAGCGC	AACGCAATTA	ATGREGETTA	GCTCACTCAT	TAGGCACCCC	AGGCTTTACA	CTTTATGCTT
1921	CCGGCTCGTA	TGTTGTGTGG	AATTGTGAGC	GGATAACAAT	TTCACACAGG	AAACAGCTAT	GACCATGATT
					KpnI	~	KhoI
					222222		******
3991	ACGCCAAGCG	CGCAATTAAC	CCTCACTAAA	GGGAACAAAA	GCTGGGTACC	ටටටටටටට	TCGAGGTCAT
4061	TCATATGCTT	GAGAAGAGAG	TCGGGATAGT	CCARATTAR	ACAAAGGTAA	GATTACCTGG	TCAAAAGTGA
131	AAACATCAGT	TAAAAGGTGG	TATAAGTAAA	ATATCGGTAA	TAAAAGGTGG	CCCAAAGTGA	AATTTACTCT
4201	TTTCTACTAT	TATAAAATT	GAGGATGTTT	TGTCGGTACT	TTGATACGTC	ATTTTTGTAT	GAATTGGTTT
1271	TTAAGTTTAT	TCGCGATTTG	GAAATGCATA	TCTGTATTTG	AGTCGGTTTT	TAAGTICGIT	GCTTTTGTAA
1341	ATACAGAGGG	ATTTGTATAA	GAAATATCTT	TAAAAAACCC	ATATGCTAAT	TTGACATAAT	TTTTGAGAAA
		ECORI					
1		1 1 1 1	1				
4411	AATATATATT	CAGGCGAATT	CCACAATGAA	CAATAATAAG	ATTAAAATAG	CTTGCCCCCG	TTGCAGCGAT
4481	GGGTATTTTT	TCTAGTAAA	TAAAAGATAA	ACTTAGACTC	AAAACATTTA	CAAAAACAAC	CCCTAAAGTC
4551	CTAAAGCCCA	AAGTGCTATG	CACGATCCAT	AGCAAGCCCA	GCCCAACCCA	ACCCAACCCA	ACCCACCCCA
4621	GTGCAGCCAA	CTGGCAAATA	GTCTCCACCC	CCGGCACTAT	CACCGTGAGT	TGTCCGCACC	ACCECACGTC
4691	TCGCAGCCAA	ABABABABAB	Agaaagaaa	AAAAGAAAA	GAAAAACAGC	AGGTGGGTCC	GGGTCGTGGG

## FIG.\_37D

GGCCGGAAAA GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCTTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCC TCTCCTCCCA TCCCCCCAAC CCTACCACCA CCACCACAC CACCTCCTCC CCCCTCGCTG CCGAACGACG AGCTCCTCCC CCTCCCCCT CCGCCCCCC CGGTAACCAC CCCGCCCCTC TCCTCTTTCT TTCTCCGTTT TTTTTTCGT CTCGGTCTCG ATCTTGCC TGGTAACTAC GGGTGGGCGA GAGCGGCTTC GTCGCCCAGA TCGGTGCGCG GGAGGGGCGG GATCTCGCGG CTGGCGTCTC BAMBI	CGGGCGTGAG TCGGCCCGGA TCTCGCGGG GAATGGGGGCT CTCGGATGTA GATCTTCTTT CTTCTTCTT TTTGTGGTAG AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTCTT TTCATGATTT GTGACAAATG
TCCGCTTCCA CCTACCACCA CCGCCGCCGC ATCTTTGGCC GATCTTCGCGG	
CGGCCCTCCC TCCCCCCAAC CCCTCCCCT CTCGGTCTCG GGAGGGGCGG	CTCGGATGTA AGTTTTTCTT
CGACGAGGCC TCTCCTCCCA AGCTCCTCCC TTTTTTTCGT TCGGTGCGCG	Gaatggggct Gttcatcggt
TCGCGAGCAG TACCCCCCC CCGGACGACG TTCTCCGTTT GTCGCCCAGA	TCCTCGCGGG CCTCAGCATT TTGTAGC
GCGAGGAGGA ACTATATACA CCCCTCGCTG TCCTCTTTCT GAGCGGCTTC	TCGGCCCGGA TCCC AATTTGAATC CCTC CGGAGCTTTT TTGI
GGCCGGAAAA CCCCATCGCC CACCTCCTCC CCCGCCCCTC GGGTGGGCGA	CGGGCGTGAG TTTGTGGTAG CAGCCTCGTG
4761 4831 4901 5041	5111 5181 5251

## FIG. 37E



**SUBSTITUTE SHEET (RULE 26)** 

Z

TICCGIGGCA CIGGIAGIGA TACGAAICIA CAACGGTTGT GAAGTACACG CAGCIGICIG CGACATACGA CAACATCCGC CIGTACACCI ICGGCGAACC GCGCAGCGGC AATCAGGCCI AGGITAGCCA O Gregaratra ratregares srcreeree assaccaast coasteserr srcaaacase GGCGCCTCCC TGGCGGCACT O SILAA Ŋ V K pq O H CTAACTACAC CCTCACGCCT TTCGACACCC TACCACAATG ρι 田の日 Ø e e **∢** ∪ ſ×ι CCACKCCCTC GCAAAGAAT AATCACCGTC H X О П BamHI K T Ø GTATCCGGAC TACGCGCTGA CCGTGACCGG A & ∀ YAL GATCCTCCGC GACGACAGCA ₩ છ Η CAACTCGATA . Y Н U

141

71

211

281

ď

E+

Þ

CCAGAACACA TITIGICTGCA CTGGGGATGA AGIGCAGTGC IGTGAGGCCC TGCCAACGAC GGCATCCCAA ACCTGCCCCC GGTGGAGCAG GGGTACGCCC ATGGCGGTGT AGAGTACTGC × TCGCGTCGTA CATGAACGAT GCCTTCCAAG CCTCGAGCCC AGATACGACG CAGTATTTCC MCOI Œ O > × ₩ 6 U A 년 . 인 Д **Q**I Ŋ 덕 Ø 4 r L 4 Z AGCGTTGATC CTTACAGCGC A E H G Ħ ρι Z 4 351 421 491

aggecggaca gggtgtgaat aatgegcaca cgacttattt tgggatgaeg ageggageet gtacatggtg GCGTAACCAC TTAACATGTA TIGAAICCIG AGAGGGGCC TITCTTAAGA CATGTAATAA GATGTCCTGG GGCAATAAG TTACGTTAAG Ø CGAGTGTACC AGGAAAGATG AGCAGATCGT TCAAACATTT TTCTGTTGAA H H ATCATATAAT N A H TCAGCCTCCC TGCGATGATT GCTGTAAAGA Z > O ATCAGTCATT FIGCCGGTCT TGAAGGATGA Ö 631 701 561 771

GABAACAAAA TATAGCGCGC AAACTAGGAT AAATTATCGC GCGCGGTGTC ATCTATGTTA CTAGATCGAT X 911

1222221

TATACATTTA ATACGCGATA

GTCCCGCAAT

TTAITITATGA GATGGGTTTT TATGATTAGA

ATGCATGACG

Hindiri

CTCACTGGCC	GCAGCCTGAA	CAGCGTGACC	ACGTTCGCCG	GGCACCTCGA	TTTTCGCCCT	AACCCTATCT	AGCTGATTTA	CGGGGAAATG	GACAATAACC	CGCCCTTATT	AAAGATGCTG	TTGAGAGTTT	ATTATCCCGT	GAGTACTCAC	TAACCATGAG	TTTTTGCAC	CCAAACGACG	AACTACTTAC	TCTGCGCTCG	GGTATCATTG	AGGCAACTAT	GTCAGACCAA	GTGAAGATCC	ACCCCGTAGA	AAAAAAACCA	ACTGGCTTCA	AGAACTCTGT	TAAGTCGTGT	GGGGGTTCGT	Tatgagaaag	AGGAGAGCGC	CTCTGACTTG	CGGCCTTTTT
ATTACGCGCG	CAACAGTTGC	TGGTTACGCG	CTTTCTCGCC	AGTGCTTTAC	GATAGACGGT	AACAACACTC	TTAAAAATG	TGGCACTTT	CCGCTCATGA	ATTTCCGTGT	GGTGAAAGTA	GGTAAGATCC	GTGGCGCGGT	TGACTTGGTT	AGTGCTGCCA	AGCTAACCGC	TGAAGCCATA	TTAACTGGCG	CAGGACCACT	TEGGTCTCGC	ACGGGGGAGTC	ATTGGTAACT	AAGGATCTAG	TGAGCGTCAG	GCTTGCAAAC	TCCGAAGGTA	CACCACTTCA	CCAGTGGCGA	GGGCTGAACG	CAGCGTGAGC	GGGTCGGAAC	GTTTCGCCAC	GCCAGCAACG
AGTGAGTCGT	TCGCCCTTCC	GCGGGTGTGG	TCTTCCCTTC	GTTCCGATT	CCATCGCCCT	TCCAAACTGG	GGCCTATTGG	ACAATTTAGG	AAATATGTAT	AGTATTCAAC	CAGAAACGCT	TCTCAACAGC	GTTCTGCTAT	ATTCTCAGAA	AGAATTATGC	GGACCGAAGG	CGGAGCTGAA	GCGCAAACTA	GATAAAGTTG	CCGGTGAGCG	TATCTACACG	CTGATTAAGC	TTTAATTTAA	TTCGTTCCAC	GTAATCTGCT	CAACTCTTTT	GTAGTTAGGC	Greecrecre	CGCAGCGGTC	GAGATACCTA	GTAAGCGGCA	GTCCTGTCGG	ATGGAAAAAC
TTCCCCCTAT	CCCGCACCGA	ATTAAGCGCG	CCTTTCGCTT	TCCCTTTAGG	ACGTAGTGGG	GGACTCTTGT	TGCCGATTTC	ATTAACGCTT	AAATACATTC	GAAGAGTATG	TTTGCTCACC	TCGAACTGGA	CACTTTTAAA	CGCATACACT	TGACAGTAAG	AACGATCGGA	CGTTGGGAAC	CAACAACGTT	GATGGAGGCG	AAATCTGGAG	GTATCGTAGT	AGGTGCCTCA	AAACTTCATT	AACGTGAGTT	TTTTCIGCCC	CAAGAGCTAC	TAGTGTAGCC	CCTGTTACCA	CCGGATAAGG	ACACCGAACT	CAGGTATCCG	TATCTTTATA	GGCGGAGCCT
GGAGCTCCAA	AGCGAAGAGG	GTAGCGGCGC	AGCGCCCGCT	AATCGGGGGC	GTGATGGTTC	CTTTAATAGT	TAAGGGATTT	TTAACAAAT	TTATTTTCT	ATTGAAAAG	CCTTCCTGTT	GTGGGTTACA	CAATGATGAG	ACTCGGTCGC	ACGGATGGCA	TACTTCTGAC	TCGCCTTGAT	GTAGCAATGG	TAATAGACTG	TATTGCTGAT	AAGCCCTCCC	TCGCTGAGAT	GATTGATTTA	AAAATCCCTT	GAGATCCTTT	TTTGCCGGAT	ACTGTCCTTC	CTCTGCTAAT	ACGATAGTTA	CGAACGACCT	GAAAGGCGGA	AAACGCCTGG	TCGTCAGGGG
AGCGGCCGGT	CTGGCGTAAT	GACGCGCCT	CCAGCGCCCT	TCAAGCTCTA	CTTGATTAGG	AGTCCACGTT	TTTTGATTTA	AACGCGAATT	CCCTATTGT	CTTCAATAAT	CGGCATTTG	GGGTGCACGA	GAACGTTTTC	GGCAAGAGCA	AAAGCATCTT	GCGCCCAACT	ATCATGTAAC	CACGATGCCT	CGGCAACAAT	CTGGCTGGTT	GCCAGATGGT	AATAGACAGA	ATATACTTTA	TCTCATGACC	GGATCTTCTT	CGGTGGTTTG	GATACCAAAT	ACATACCTCG	TGGACTCAAG	CAGCTTGGAG	CCCGAAGGGA	TTCCAGGGG	TTTGTGATGC
AAGCTTCTAG	CTTTCGCCAG	TGGCGAATGG	GCTACACTTG	GCTTTCCCCG	CCCCAAAAA	TTGACGTTGG	CGGTCTATTC	ACAAAAATTT	TGCGCGGAAC	CTGATAAATG	CCCTTTTTG	AAGATCAGTT	TCGCCCCGAA	ATTGACGCCG	CAGTCACAGA	TGATAACACT	AACATGGGGG	AGCGTGACAC	TCTAGCTTCC	GCCCTTCCGG	CAGCACTGGG	GGATGAACGA	GTTTACTCAT	TTTTTGATAA	AAAGATCAAA	CCGCTACCAG	GCAGAGCGCA	AGCACCGCCT	CTTACCGGGT	GCACACAGCC	CGCCACGCTT	ACGAGGGAGC	AGCGTCGATT
981	1121	1191	1261	1331	1401	1471	1541	1611	1681	1751	1821	1891	1961	2031	2101	2171	2241	2311	2381	2451	2521	2591	2661	2731	2801	2871	2941	3011	3081	3151	3221	3291	3361

# 7G.\_38C

TITCITITICA IGAITITGIGA CAAAIGCAGC CICGIGGGGA GCITITITIGI AGCAAGCITA	GCTITITIO	CTCGTGCGGA	CAAATGCAGC	TGATTTGTGA	TTTCTTTTCA	ATCGGTAGTT	5041
GGGGCTCTCG CATGTAGATC TTCTTTTT CTTCTTTTG TGGTAGAATT TGAATCCCTC AGCATTGTTC HIDGIII	TGAATCCCTC	tggtagaatt	CTTCTTTTG	TTCTTTCTTT	Gatgtagatc	GGGGCTCTCG	4971
		Belli	i				
TACCCCCAGAA GAGGAATC TCGCGGCTGG CGTCTCCCGG CGTGAGTCGG CCCGGAATCT CGCGGAAAT	CCCGGATCCT	CGTGAGTCGG	CGTCTCCGGG	TCGCGGCTGG	GGGCGGGATC	TGCGCGGGAG	4901
Bamil	<b>A</b>						
CCCAGATCGG	GGCTTCGTCG	GGGCGAGAGC	TAGTTTGGGT	Treecctres	GTCTCGATCT	TITCGICICG	4831
	CTTTCTTTCT	CCCCTCTCCT	AACCACCCG	にのここのここのここの	CCCCTCCGC	CCICCCCCC	4761
ACGACGAGCT	TCGCTGCCGG	TCCTCCCCC	CACCACCACC	CCACCACCAC	CCCAACCCTA	CICCCAICCC	4691
	TATACATACC	ATCGCCACTA	AACGCCCCCC	CTTCCAAAGA	CCTCCCTCCG	GAGGCCCGGC	4621
GAGCAGCGAC	GGAGGATCGC	GGAAAAGCGA	CGTGGGGGCC	GGGTCCGGGT	AACAGCAGGT	GAAAAAGAAA	4551
	AAAAAAAAA	AGCCAAAAA	CACGICICGC	CGCACCACCG	GTGAGTTGTC	CACTATCACC	4481
	CAAATAGTCT	AGCCAACTGG	ACCCCAGTGC	AACCCAACCC	AACCCAACCC	AGCCCAGCCC	4411
AGATAGATT	GLAPARIAGO	ATTITITE	AGCGATGGGT	いったこうこうこう	AAATAGCIIG	AATAAGATTA	4271
	CGAATTCCAC	TATATTCAGG	GAGAAAATA	CATAATTTT	GCTAATTTGA	AAACCCATAT	4201
112222	2						
ECORI							
TATCTTTAAA	GTATAAGAAA	AGAGGGATTT	TTGTAAATAC	TTCGTTGCTT	GGTTTTTAAG	TATTTGAGTC	4131
	GATTTGGAAA	GTTTATTCGC	TGGTTTTTAA	TTGTATGAAT	TACGICALIT	GGTACTTTGA	4061
		ALCHOITANA TACTATUTATA	というない こうかん はんしょう	ALCIGOTORS PHESS PHE	T TWO WE TO DO	ARTEMBER AND OF	2 4 6 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
	AGAGAGTCGG	ATGCTTGAGA	GGTCATTCAT	CCCCCTCGA	GGTACCGGGC	ACAAAAGCTG	3851
		222222	1 2 2	1 1 1			
		XhoI	H	KpnI			
ACTAAAGGGA	ATTAACCCTC	CAAGCGCGCA	AGCTATGACC ATGATTACGC	AGCTATGACC	CACAGGAAAC	AACAATTTCA	3781
_	GTGTGGAATT	CTCGTATGTT	ATGCTTCCGG	TTTACACTTT	CACCCAGGC	ACTCATTAGG	3711
GAGTTAGCTC	CAATTAATGT	GAGCGCAACG	AGCGGGCAGT	CCGACTGGAA	GACAGGTTTC	CAGCTGGCAC	3641
		TCTCCCCCC		CGCCCAATAC	AGCGGAAGAG	TGAGCGAGGA	3571
AGCGAGTCAG	TATCCCCTGA	Trrccrecer gradeceaaac	TCACATGTTC	GCCITITIGE GGCCITITIGC ICACATGITC ITICCIGGGI IAICCCCGG CCGCCTTTGA GTGAGCTGAT ACCGCTCGCC GCAGCCGAAC GACCGAGCGC	CCGCCTTTGA	ACCGTTCCTG	3501

#### -1G.\_38D

1184

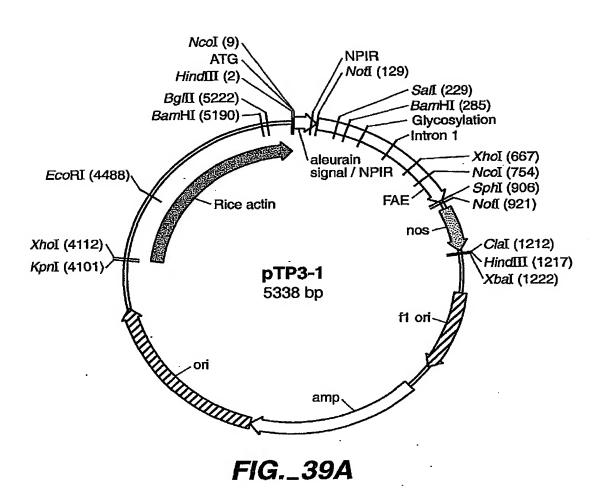
S A K H V L A V V V T A G H A L A A B CCCGC AACACGCCT TAGCACGCCTC S E D L Y S R L V E M A T I S Q A A Y CCGAAG ACCTCTACAAG ACCTCTACAG CCGTTTAGTC GAAATGGCCA CTATCTCCCA AGCTGCCTAC

I I K G E K I Y N IATT ATCAAGGAG AGAAAATTTA CAAIT

5251

5181

FIG. 38



NGOI

Hindrin

Н

AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG H н

Noti

NPIR

GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC đ GACCGCGCGG GCCCGTCACC A D S N P I R GCCGACTCCG Н S 8 F CTCCTCCTTC 6 · A S S T

Salı

141

1

TTCTCAAACT GACATTAACG O W GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA 妇 Ō Н Ø ບ H 211

BamHI

GTGATACGAA σχ GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA Ø ש . 14 [h > 回 ω ⋈ R D D WHL 281

Glycosylation

1

COCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG cerceacace receraacaa caereaerae GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA 臼 ບ CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA Ø Z U O . 단 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC R L Ē щ ۴ H ₩ V 83 A A TCTACAACTC ש 561 351 421 491

CGTTACCCAA ACCGATCGCC

TCCAATTCGC

2222222

H Þ œ GACCCAGTAT OI H GCCCAGATAC H Ω Ŋ CAAGCCTCGA Ø CGATGCCTTC **F4** 4 A CGTACATGAA Z GCCTTCGCGT

GTGTAGAGTA 国 U Þ ບ Ö GCCCATGGCG 11111111 Ö NCOL Ħ M A GCAGGGGTAC Ö Ø H O CCAAACCTGC CCCGGTGGA Д H μ O z ಭ CGACGGCATC Þ Ø ρį A CTCATGCCAA Z 4 I 8

GTGCTGTGAG TGCACTGGGG ATGAAGTGCA CACATTTGTC

Sphi

GATCCTTACA GCGCCCAGAA CTGGAGCGTT

701

631

771

U 4 GACGAGCGGC Ö Ø H ATTTTGGGAT Ħ O ÎΞ × CACACGACTT Н H H GAATAATGCG ⋖ Z Z GACAGGGTGT > Ø O ø GCCCAGGGCG ש O

四四

Noti

Ø Н 22222222

TTGAATTACG ATCGCGCGCG TTAGAGTCCC CATTTGGCAA AAAGAAGCAG ATCGTTCAAA ATAATTTCTG AGGATAAATT GITTITATEA TGATTATCAT CGCGCAAACT TATGAGATGG ACCACTGAAG GATGAGCTGT CGATAGAAA CAAAATATAG TGACGTTATT GGTCTTGCGA ATGTAATGCA TCCTGTTGCC ATTTAATACG TAAGATTGAA AATAATTAAC GGCCGCGGAA GCAATTATAC TTAAGCATGT GGCCGGTCGC TAAAGTTTCT 1051 911 981

HindIII

1222222

CCGGTGGAGC TCTAGAGCGG TCGATAAGCT TGTTACTAGA

AAAACCCTGG AGAGGCCCGC GGCGCATTAA GTAATAGCGA GCCCTGTAGC CGTGACTGGG AATGGGACGC GCCAGCTGGC TTTACAACGT CTGAATGGCG TCCCCCTTTC TGCCCGTCGT GCGCGCTCAC TTGCAGCACA GTTGCGCAGC CTTAATCGCC CTTCCCAACA GTGTCATCTA GTCGTATTAC 1331 1401 1261

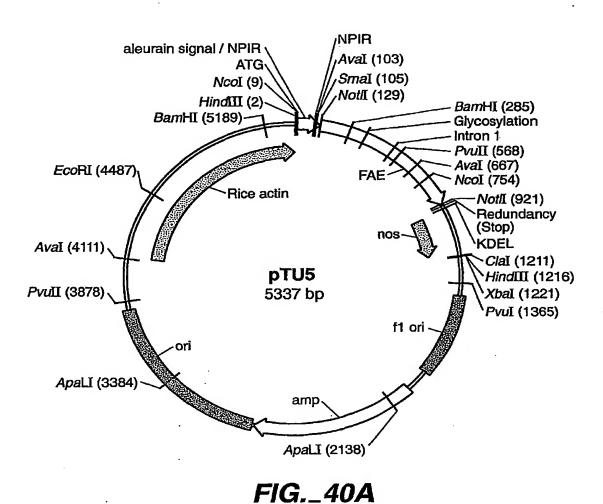
841

CTTC	TICC	CATC	CCAA	GCCT	CAAT	AATA	GTAT	AGAA	CTCA	TTCT	TICI	GAAT	GACC	GGAG	CGCA	ATAA	CGGT	ATCT	TGAT	TTAA	TCGT	TAAT	AACT	TAGT	TGGC	GCAG	AGAT	TAAG	TCCT	TGGA	TTCC	CAGC	CHCC	P GCG
_	_	_	_	ATTTCGGCCT	CGCTTACAAT	CATTCAAATA	GTATGAGTAT	TCACCCAGAA	CTGGATCTCA	TTAAAGTTCT	ACACTATICI	GTAAGAGAAT	TCGGAGGACC	GGAACCGGAG	ACGITGCGCA	AGGCGGATAA	TGGAGCCGGT	GIAGITATCI	CCTCACTGAT	TCATITITAN	CAGTITICGI	TGCGCGTAAT	_			-:	GAACTGAGAT	ATCCGGTAAG			•	_	Ξ.	GCAGTGAGCG
CCGCTCCTTT	GGGGCTCCT	GGTTCACGTA	ATAGTGGACT	GATTTTGCCG	AAAAAAAAA	TTTCTAAATA	AAAAGGAAGA	CTGTTTTGC	TTACATCGAA	ATGAGCACTT	GTCGCCGCAT	TGGCATGACA	CTGACAACGA	TTGATCGTTG	AATGGCAACA	GACTGGATGG	CTGATAAATC	CICCCGIAIC	GAGATAGGTG	ATTTAAAACT	CCCTTAACGT	CCTTTTTTC	CGGATCAAGA	CCTTCTAGTG	CTAATCCTGT	AGTTACCGGA	GACCTACACC	GCGGACAGGT	CCTGGTATCT	AGGGGGGCGG	TTTGCTCACA	CTGATACCGC	AATACGCAAA	TGGANAGCGG
GCCCTAGCGC	CTCTAAATCG	TTAGGGTGAT	ACGUTCTTTA	ATTTATAAGG	GAATTTTAAC	TTTTGTTTATT	ATAATATTGA	TTTTGCCTTC	CACGAGTGGG	TTTTCCAATG	GAGCAACTCG	ATCTTACGGA	CAACTTACTT	GTAACTCGCC	TGCCTGTAGC	ACAALTAATA	TGGTTTATTG	ATGGTAAGCC	ACAGATCGCT	CTTTAGATTG	TGACCAAAAT	TTCTTGAGAT	GTTTGTTTGC	CAAATACTGT	CCICGCICIG	TCAAGACGAT	TGGAGCGAAC	AGGGAGAAAG	GGGGGAAACG	GATGCTCGTC	TIGCIGGCCI	TTTGAGTGAG	AAGAGCGCCC	GTTTCCCGAC
ACTTGCCAGC	CCCCGTCAAG	AAAAACTTGA	GTTGGAGTCC	TATTCTTTTG	AATTTAACGC	GGAACCCCTA	AAATGCTTCA	TTTTGCGGCA	CAGTTGGGTG	CCGAAGAACG	CGCCGGGCAA	ACAGAAAAGC	ACACTGCGGC	GGGGGATCAT	GACACCACGA	CTTCCCGGCA	TCCGGCTGGC	CTGGGGCCAG	AACGAAATAG	CTCATATATA	GATAATCTCA	TCAAAGGATC	ACCAGCGGTG	GCGCAGATAC	CGCCTACATA	CGGGTTGGAC	CAGCCCAGCT	CGCTTCCCGA	GGAGCTTCCA	CGATTTTTGT	TCCTGGCCTT	TATTACCGCC	GAGGAAGCGG	GGCACGACAG
TGACCGCTAC	CGCCGGCTTT	CTCGACCCCA	GCCCTTTGAC	TATCTCGGTC	ATTTAACAAA	AAATGTGCGC	TAACCCTGAT	TTATTCCCTT	TGCTGAAGAT	AGTTTTCGCC	CCCGTATTGA	CTCACCAGTC	ATGAGTGATA	TGCACAACAT	CGACGAGCGT	CTTACTCTAG	GCTCGGCCCT	CATTGCAGCA	ACTATGGATG	ACCAAGTTTA	GATCCTTTTT	GTAGAAAGA	AACCACCGCT	CTTCAGCAGA	TCTGTAGCAC	CGTGTCTTAC	TTCGTGCACA	GAZAGCGCCA	AGCGCACGAG	ACTIGAGCGI	TTTTTACGGT	TGGATAACCG	GTCAGTGAGC	TAATGCAGCT
ACGCGCAGCG	TCGCCACGTT	TTTACGGCAC	ACGGTTTTTC	CACTCAACCC	AAATGAGCTG	CTTTTCGGGG	CATGAGACAA	CGTGTCGCCC	AAGTAAAGA	GATCCTTGAG	GCGGTATTAT	TGGTTGAGTA	TGCCATAACC	ACCGCTTTTT	CCATACCAAA	TGGCGAACTA	CCACTTCTGC	CTCGCGGTAT	GAGTCAGGCA	TAACTGTCAG	TCTAGGTGAA	GTCAGACCCC	CAAACAAAAA	AGGTAACTGG	CTTCAAGAAC	GGCGATAAGT	GAACGGGGGG	TGAGCTATGA	GGAACAGGAG	GCCACCTCTG	CAACGCGGCC	CCTGATTCTG	AGCGCAGCGA	GCCGATTCAT
TGTGGTGGTT	CCTTCCTTTC	GATTTAGTGC	GCCCTGATAG	ACTGGAACAA	ATTGGTTAAA	TTAGGTGGCA	TGTATCCGCT	TCAACATTTC	ACGCTGGTGA	ACAGCGGTAA	GCTATGTGGC	CAGAATGACT	TATGCAGTGC	GAAGGAGCTA	CTGAATGAAG	AACTATTAAC	AGTTGCAGGA	GAGCGTGGGT	ACACGACGGG	TAAGCATTGG	TTTAAAAGGA	TCCACTGAGC	CIGCIGCITG	CTTTTTCCGA	TAGGCCACCA	TGCTGCCAGT	CGGTCGGGCT	ACCTACAGCG	CGGCAGGGTC	GTCGGGTTTC	AAAACGCCAG	TGCGTTATCC	CGAACGACCG	CCGCGCGTTG
1471	1541	1611	1681	1751	1821	1891	1961	2031	2101	2171	2241	2311	2381	2451	2521	2591	2661	2731	2801	2871	2941	3011	3081	3151	3221	3291	3361	3431	3501	3571	3641	3711	3781	3851

# FIG.\_39D

CAACGCAATT AATGIGACTT AGCTCACTCA TTAGGCACCC CAGGCTTAC ACTTTATGCT ATGTGTGTG GAALTGTGAG GAACAGCTA TGACCATGAT TTAGACACA TTTCACACAG GAACAGCTA TGACCATTAA AGGGAACAAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTCGAGGTCA TTAGAAAAGTTA AGAATTCGTA ATGAAAAGTA AGAATTCGTA TTAGACATTT GGAAATTCTT TTGACGATT TTGACGATTT TTGACGATTTTTGAGAA TGAATTTGTTA AGAATTCTT TTAAAAAGTTA ATGACTTTTTTGAGAA TTTTTTTGAGAA TTTTTTTTTT	
CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ATGTTGTGA GAATTGTGAG CGGATAACAA TTTCACACACG GAACCAGCTA  CCGCAATTAA CCCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCCC TGAGAAGGTG GTATAAGTAA AATATCGGTA ATAAAAGGTG GCCCAAGTG TTATAAAAGGTG GTATAAGTAA AATATCGGTA ATAAAAGGTG GCCCAAGTG TTATAAAAATT GGAAATGTT TTGTCGCTAC TTTGACTTTTTTTTTT	
	CCCTACCAC CCCTACCACCAC GATCTTTGGC GGATCTCGCG BGILI AGATCTTCTT TTTCATGATT
	GCGACGAGGGCCCTCC CTCTCCTCCC ATCCCCCCAA GAGCTCCTCC CCCCTCCCCCAA TTTTTTTTCG TCTCGGTCTC ATCGGTGCGC GGGAGGGGCG GGAATGGGGC TCTCGGATGT TGTTCATCGG TAGTTTTTCT
	AGCETAGRAS ATCACERAGEA GUGACUAGAGE CUGGECUCTOC CACTATATAC ATACCCCCCC CTCTCCTCCC ATCCCCCCAA CCCCCTCTTTC TTTCTCCGTT TTTTTTTCG TCTCGGTCTC AGAGCGGCTT CGTCGCCCAA ATCGGTGCGC GGGAGGGGCG Bambi CACACCCCGA ATCGTCGCG GGAATGGGGC TCTCGGATGT GTCGGCCCGG ATCCTCGCGG GGAATGGGGC TCTCGGATGT GAATTTGAAT CCTCAGCAT TGTTCATCGG TAGTTTTTCT GCGGAGCTTT TTTGTAGC
	AGCGAGGAGG ATCGCGAGGCA CACTATATAC ATACCCCCC CCCCTCGTT GCCGACGAC AGAGCGGCTT CGTCGCCCAG BAMBI GTCGGCCCGG ATCCTCGCGG GAATTTGAAT CCCTCAGCAT GCGGAGCTTT TTTGTAGC
ee eeeee	AGCGAGGAGG ATCGCGAGGCCCCCCCCCCCCCCCCCCGACGCCCCTTTCTTCTCCCCCCCC
33 39 39 39 39 39 39 39 39 39 39 39 39 3	6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6

## FIG.\_39E



SUBSTITUTE SHEET (RULE 26)

WO 02/068666

76/154

	HindIII No	οĪ			
	~~~~~~~~			•	
1		TGGCCCACGC			
	TTCGAATGGT	ACCGGGTGCG	GGCGCAGGAG	GAGGACCGCG	AGCGGCACGA
51	GGCCACGGCC	GCCGTCGCCG	TCGCCTCCTC	CTCCTCCTTC	GCCGACTCCA
	CCGGTGCCGG	CGGCAGCGGC	AGCGGAGGAG	GAGGAGGAAG	CGGCTGAGGT
	SmaI				
	AvaI		Not	:I	
	~~~~			~~~~	
101		GCCCGTCACC			
	TGGGCCCGGC	CGGGCAGTGG	CTGGCGCGCC	GGCGGAGGTG	CGTCCCGTAG
151	TCCGAAGACC	TCTACAGCCG	TTTAGTCGAA	ATGGCCACTA	TCTCCCAAGC
		AGATGTCGGC			
201		GACCTGTGCA			
	ACGGATGCGG	CTGGACACGT	TGTAAGGCAG	CTGATAATAG	TTCCCTCTCT
				BamHI	
				~~~~~	
251		TTCTCAAACT			
	TITAAATGTT	aagagtttga	CIGIAATIGC	CTACCTAGGA	GGCGC1GC1G
301	AGCAGCAAAG	AAATAATCAC	CGTCTTCCGT	GGCACTGGTA	GTGATACGAA
	TCGTCGTTTC	TTTATTAGTG	GCAGAAGGCA	CCGTGACCAT	CACTATGCTT
254	mama an a ama	GATACTAACT		acammaca a	X CCCTTX CCX C
351	AGATGTTGAG	CTATGATTGA	TOTOGORAGTO	CGGAAAGCTG	TGGGATGGTG
			10100011010		
401		TTGTGAAGTA			
	TTACGTTGCC	AACACTTCAT	GTGCCACCTA	TAATATAACC	TACCCAGAGG
451	GTCCAGGACC	AAGTCGAGTC	CCTTGTCAAA	CAGCAGGTTA	GCCAGTATCC
- L		TTCAGCTCAG			
501		CTGACCGTGA			
	CCTGATGCGC	GACTGGCACT	GGCCGGTGMG	GGAGCCGCGG	AGGGACCGCC
		PvuII			
		~~~~			
551		CGCCCAGCTG			
	GTGAGTGACG	GCGGGTCGAC	AGACGCTGTA	TGCTGTTGTA	GGCGGACATG
601	ACCTTCGGCG	AACCGCGCAG	CGGCAATCAG	GCCTTCGCGT	CGTACATGAA
-		TTGGCGCGTC			
		AvaI			
651	CGZ diCCCdrinc	CAAGCCTCGA		<b>ር</b> ልሮርር አርጥልጥ	<b>ተጥሮሮሮ</b> ርርጥሮ እ
<b>5</b> 51		GTTCGGAGCT			

FIG.\_40B

77/154

701				CCCCGGTGGA GGGGCCACCT	
	Ncol				
751				GATCCTTACA CTAGGAATGT	
801				GTGCTGTGAG CACGACACTC	
851				ATTTTGGGAT TAAAACCCTA	
			NotI		
901	GCATGCACCT CGTACGTGGA	GGCCGGTCGC	GGCCGCGGAA	CCACTGAAGG GGTGACTTCC	
951				AAAGTTTCTT TTTCAAAGAA	
1001	CCTGTTGCCG	GTCTTGCGAT	GATTATCATA	TAATTTCTGT	TGAATTACGT
	GGACAACGGC	CAGAACGCTA	CTAATAGTAT	ATTAAAGACA	ACTTAATGCA
1051				GACGTTATTT CTGCAATAAA	
1101				TTTAATACGC AAATTATGCG	
1151				TCGCGCGCGG AGCGCGCGCC	
			XbaI		
	G.	 lar HindIII	ynnann T		
		rar utnott			
1201				CGGTGGAGCT GCCACCTCGA	
1251				GGCCGTCGTT CCGGCAGCAA	
1301				TTAATCGCCT AATTAGCGGA	
		PvuII			
1351	CCCCCTTTCG	CCAGCTGGCG	TANTAGCGAA	GAGGCCCGCA	CCGATCGCCC
				CTCCGGGCGT	
1401	TTCCCAACAG AAGGGTTGTC			ATGGGACGCG TACCCTGCGC	

FIG.\_40C

78/154

1451				CGCGCAGCGT GCGCGTCGCA	
1501				GCTTTCTTCC	
	GAACGGTCGC	GGGATCGCGG	GCGAGGAAAG	CGAAAGAAGG	GAAGGAAAGA
1551				TCTAAATCGG AGATTTAGCC	
1601	TAGGGTTCCG ATCCCAAGGC	ATTTAGTGCT TAAATCACGA	TTACGGCACC AATGCCGTGG	TCGACCCCAA AGCTGGGGTT	AAAACTTGAT TTTTGAACTA
1651				CCCTGATAGA GGGACTATCT	
1701				TAGTGGACTC ATCACCTGAG	
1751	CTGGAACAAC	ACTCAACCCT	ATCTCGGTCT	ATTCTTTTGA TAAGAAAACT	TTTATAAGGG
1801	ATTTTGCCGA TAAAACGGCT	TTTCGGCCTA AAAGCCGGAT	TTGGTTAAAA AACCAATTTT	AATGAGCTGA TTACTCGACT	TTTAACAAAA AAATTGTTTT
1851	ATTTAACGCG	AATTTTAACA	AAATATTAAC	GCTTACAATT	TAGGTGGCAC
	TAAATTGCGC.	TTAAAATTGT	TTTATAATTG	CGAATGTTAA	ATCCACCGTG
1901	TTTTCGGGGA	AATGTGCGCG	GAACCCCTAT	TTGTTTATTT	TTCTAAATAC
	AAAAGCCCCT	TTACACGCGC	CTTGGGGATA	AACAAATAAA	AAGATTTATG
1951	ATTCAAATAT	GTATCCGCTC	ATGAGACAAT	AACCCTGATA	AATGCTTCAA
	TAAGTTTATA	CATAGGCGAG	TACTCTGTTA	TTGGGACTAT	TTACGAAGTT
2001				CAACATTTCC	
	ATTATAACTT	TTTCCTTCTC	ATACTCATAA	GTTGTAAAGG	CACAGCGGGA
2051	TATTCCCTTT	TTTGCGGCAT	TTTGCCTTCC	TGTTTTTGCT	CACCCAGAAA
	ATAAGGGAAA	AAACGCCGTA	AAACGGAAGG	ACAAAAACGA	GTGGGTCTTT
				Apa	LI ~~~
2101	CGCTGGTGAA	AGTAAAAGAT	GCTGAAGATC	AGTTGGGTGC	ACGAGTGGGT
	GCGACCACTT	TCATTTTCTA	CGACTTCTAG	TCAACCCACG	TGCTCACCCA
2151	TACATCGAAC	TGGATCTCAA	CAGCGGTAAG	ATCCTTGAGA	GTTTTCGCCC
•	ATGTAGCTTG	ACCTAGAGTT	GTCGCCATTC	TAGGAACTCT	CAAAAGCGGG
2201	CGAAGAACGT	TTTCCAATGA	TGAGCACTTT	TAAAGTTCTG	CTATGTGGCG
	GCTTCTTGCA	AAAGGTTACT	ACTCGTGAAA	ATTTCAAGAC	GATACACCGC
2251	CGGTATTATC	CCGTATTGAC	GCCGGGCAAG	AGCAACTCGG	TCGCCGCATA
	GCCATAATAG	GGCATAACTG	CGGCCCGTTC	TCGTTGAGCC	AGCGGCGTAT
2301	CACTATTCTC	AGAATGACTT	GGTTGAGTAC	TCACCAGTCA	CAGAAAAGCA
	GTGATAAGAG	TCTTACTGAA	CCAACTCATG	AGTGGTCAGT	GTCTTTTCGT
		E1	G $ADD$	(	

**FIG.\_40D** 

WO 02/068666 PCT/US01/43588

2351			TAAGAGAATT ATTCTCTTAA		
2401	TGAGTGATAA ACTCACTATT		AACTTACTTC TTGAATGAAG		
2451			GCACAACATG CGTGTTGTAC		
2501	+ +		TGAATGAAGC ACTTACTTCG		
2551			ATGGCAACAA TACCGTTGTT		
2601			TTCCCGGCAA AAGGGCCGTT		
2651			CACTTCTGCG GTGAAGACGC		
2701			GGAGCCGGTG CCTCGGCCAC		
2751			TGGTAAGCCC ACCATTCGGG		
2801			CTATGGATGA GATACCTACT		
2851			AAGCATTGGT TTCGTAACCA		
2901	TCATATATAC AGTATATATG		TTTAAAACTT AAATTTTGAA		
2951			ATAATCTCAT TATTAGAGTA		
3001			TCAGACCCCG AGTCTGGGGC		
3051	TCTTGAGATC AGAACTCTAG				
3101	ACCACCGCTA TGGTGCCGAT		TTTGTTTGCC AAACAAACGG		
3151	TTTTTCCGAA AAAAAGGCTT	GGTAACTGGC	TTCAGCAGAG	CGCAGATACC	AAATACTGTC
3201	CTTCTAGTGT	AGCCGTAGTT		TTCAAGAACT	CTGTAGCACC
3251	GCCTACATAC	CTCGCTCTGC	TAATCCTGTT	ACCAGTGGCT	GCTGCCAGTG
	COCKIGIATO	CAUCUAUACU	ATTAGGACAA	IGGICACCGA	CONCOURTENCE

#### FIG.\_40E

3301			GGGTTGGACT CCCAACCTGA		
				ApaLI	
3351	AAGGCGCAGC TTCCGCGTCG		AACGGGGGGT TTGCCCCCA		
3401			AACTGAGATA		
	***************************************		TTGACTCTAT	•	
3451		GCTTCCCGAA CGAAGGGCTT	GGGAGAAAGG CCCTCTTTCC	CGGACAGGTA GCCTGTCCAT	TCCGGTAAGC AGGCCATTCG
3501	GGCAGGGTCG				
	CCGTCCCAGC	CTTGTCCTCT	CGCGTGCTCC	CTCGAAGGTC	CCCCTTTGCG
3551	CTGGTATCTT GACCATAGAA		TCGGGTTTCG AGCCCAAAGC		
3601	GATITTTGTG	ATGCTCGTCA	GGGGGGCGGA	GCCTATGGAA	AAACGCCAGC
	CTAAAAACAC	TACGAGCAGT	CCCCCCCCT	CGGATACCTT	TTTGCGGTCG
3651	AACGCGGCCT TTGCGCCGGA		CCTGGCCTTT GGACCGGAAA		
3701			CTGATTCTGT GACTAAGACA		
3751	TTGAGTGAGC	TGATACCGCT	CGCCGCAGCC	GAACGACCGA	GCGCAGCGAG
	AACTCACTCG				
3801	TCAGTGAGCG	AGGAAGCGGA	AGAGCGCCCA	ATACGCAAAC	CGCCTCTCCC
	AGTCACTCGC	TCCTTCGCCT	TCTCGCGGGT	TATGCGTTTG	GCGGAGAGGG
			PvuII		
3851	CGCGCGTTGG				
2224					
3901	GGAAAGCGGG CCTTTCGCCC		TTGCGTTAAT		
3951	TAGGCACCCC ATCCGTGGGG	AGGCTTTACA TCCGAAATGT	CTTTATGCTT GAAATACGAA	CCGGCTCGTA GGCCGAGCAT	TGTTGTGTGG ACAACACACC
4001	AATTGTGAGC	GGATAACAAT	TTCACACAGG	AAACAGCTAT	GACCATGATT
	TTAACACTCG	CCTATTGTTA	AAGTGTGTCC	TTTGTCGATA	CTGGTACTAA
4051	ACGCCAAGCG				
	TGCGGTTCGC	GCGTTAATTG	GGAGTGATTT	CCCTTGTTTT	CGACCCATGG
		AvaI			
4101	GGGCCCCCC	TCGAGGTCAT	TCATATGCTT AGTATACGAA	GAGAAGAGAG CTCTTCTCTC	TCGGGATAGT AGCCCTATCA
•					

FIG.\_40F

4151	CCAAAATAAA	ACAAAGGTAA	GATTACCTGG	TCAAAAGTGA	AAACATCAGT
	GGTTTTATTT	TGTTTCCATT	CTAATGGACC	AGTTTTCACT	TTTGTAGTCA
4201	TAAAAGGTGG	TATAAGTAAA	ATATCGGTAA	TAAAAGGTGG	CCCAAAGTGA
	ATTTTCCACC	ATATTCATTT	TATAGCCATT	ATTTTCCACC	GGGTTTCACT
4251	AATTTACTCT	TTTCTACTAT	TATAAAAATT	GAGGATGTTT	TGTCGGTACT
	TTAAATGAGA	AAAGATGATA	ATATTTTTAA	CTCCTACAAA	ACAGCCATGA
4301	TTGATACGTC	ATTTTTGTAT	GAATTGGTTT	TTAAGTTTAT	TCGCGATTTG
	AACTATGCAG	TAAAAACATA	CTTAACCAAA	AATTCAAATA	AGCGCTAAAC
4351	GAAATGCATA	TCTGTATTTG	AGTCGGTTTT	TAAGTTCGTT	GCTTTTGTAA
	CTTTACGTAT	AGACATAAAC	TCAGCCAAAA	ATTCAAGCAA	CGAAAACATT
4401	ATACAGAGGG	ATTTGTATAA	GAAATATCTT	TAAAAAACCC	ATATGCTAAT
	TATGTCTCCC	TAAACATATT	CTTTATAGAA	ATTTTTTGGG	TATACGATTA
				EcoR	
4454		TTTTGAGAAA	NAME WAS AND A STATE OF THE STA	CACCCGAAMT	
4451	AACTGTATTA	AAAACTCTTT	TTATATATAA	GTCCGCTTAA	GGTGTTACTT
4501	CAATAATAAG	ATTAĀAATAG	CTTGCCCCCG	TTGCAGCGAT	GGGTATTTTT
	GTTATTATTC	TAATTTTATC	GAACGGGGGC	AACGTCGCTA	CCCATAAAAA
4551	TCTAGTAAAA	TAAAAGATAA	ACTTAGACTC	AAAACATTTA	CAAAAACAAC
	AGATCATTTT	ATTTTCTATT	TGAATCTGAG	TTTTGTAAAT	GTTTTTGTTG
4601	CCCTAAAGTC	CTAAAGCCCA	AAGTGCTATG	CACGATCCAT	AGCAAGCCCA
	GGGATTTCAG	GATTTCGGGT	TTCACGATAC	GTGCTAGGTA	TCGTTCGGGT
4651	GCCCAACCCA	ACCCAACCCA	ACCCACCCCA	GTGCAGCCAA	CTGGCAAATA
	CGGGTTGGGT	TGGGTTGGGT	TGGGTGGGGT	CACGTCGGTT	GACCGTTTAT
4701	GTCTCCACCC	CCGGCACTAT	CACCGTGAGT	TGTCCGCACC	ACCGCACGTC
	CAGAGGTGGG	GGCCGTGATA	GTGGCACTCA	ACAGGCGTGG	TGGCGTGCAG
4751	AGCGTCGGTT	AAAAAAAAA TTTTTTTTTT	TCTTTCTTT	TTTTCTTTT	CTTTTTGTCG
4801	AGGTGGGTCC	GGGTCGTGGG	GGCCGGAAAA	GCGAGGAGGA	TCGCGAGCAG
	TCCACCCAGG	CCCAGCACCC	CCGGCCTTTT	CGCTCCTCCT	AGCGCTCGTC
4851	CGACGAGGCC	CGGCCCTCCC	TCCGCTTCCA	AAGAAACGCC	CCCCATCGCC
	GCTGCTCCGG	GCCGGGAGGG	AGGCGAAGGT	TTCTTTGCGG	GGGGTAGCGG
4901	ACTATATACA	TACCCCCCCC	TCTCCTCCCA	TCCCCCAAC	CCTACCACCA
	TGATATATGT	ATGGGGGGGG	AGAGGAGGGT	AGGGGGGTTG	GGATGGTGGT
4951	CCACCACCAC	CACCTCCTCC	CCCCTCGCTG	CCGGACGACG	AGCTCCTCCC
	GGTGGTGGTG	GTGGAGGAGG	GGGGAGCGAC	GGCCTGCTGC	TCGAGGAGGG
5001	CCCTCCCCT GGGAGGGGA	CCGCCGCCGC	CGGTAACCAC GCCATTGGTG	CCCGCCCCTC GGGCGGGGAG	TCCTCTTTCT AGGAGAAAGA

FIG.\_40G

5051				ATCTTTGGCC	
	AAGAGGCAAA	AAAAAAAGCA	GAGCCAGAGC	TAGAAACCGG	AACCATCAAA
5101	GGGTGGGCGA	GAGCGGCTTC	GTCGCCCAGA	TCGGTGCGCG	GGAGGGGCGG
				AGCCACGCGC	
				Bar	nHI
				~~~	
5151	GATCTCGCGG	CTGGCGTCTC	CGGGCGTGAG	TCGGCCCGGA	TCCTCGCGGG
	CTAGAGCGCC	GACCGCAGAG	GCCCGCACTC	AGCCGGGCCT	AGGAGCGCCC
5201	GNATICCCCCT	CTCGGATGTA	C A de Calade Caladada	CTTTCTTCTT	TTTGTGGTAG
J201		GAGCCTACAT		GAAAGAAGAA	
5251	AATTTGAATC	CCTCAGCATT		AGTTTTTCTT	
	TTAAACTTAG	GGAGTCGTAA	CAAGTAGCCA	TCAAAAAGAA	AAGTACTAAA
5301	GTGACAAATG	CAGCCTCGTG	CGGAGCTTTT	TTGTAGC	
		GTCGGAGCAC			

FIG.\_40H

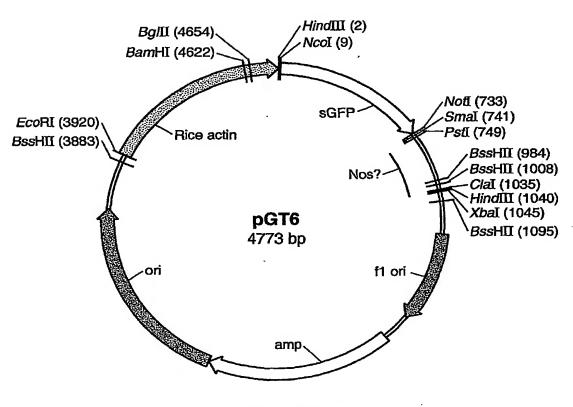


FIG.\_41A

r	٠	
1	3	
ľ	3	
t	ž	
ŕ	•	
ŀ	4	
ī.	j	
:	:	
۲	4	
q	3	
1	1	
	3	
2	3	
μ	9	
٠		

-	1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTCACCG GGGTGGTGCC CATCCTGGTC GAGCTGGACG	TGGTGAGCAA	GGGCGAGGAG	CTGTTCACCG	GGGTGGTGCC	CATCCTGGTC	GAGCTGGACG
	TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG GTAGGACCAG CTCGACCTGC	ACCACTCGTT	CCCGCTCCTC	GACAAGTGGC	CCCACCACGG	GTAGGACCAG	CTCGACCTGC

<b>,</b>	Ph.
CTGA	GACIK
GCAAG	CGTTC
LACG	ATGC
71 GCGACGIGAA CGGCCACAAG TICAGCGTGI CCGGCGAGGG CGAGGGCGAF GCCACCTACG GCAAGCTGAC	CGCTGCACIT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTCGACTG
년 달	S €
9	ည်
1860	Ď
20	ဗ
<b>366</b>	ដ្ឋ
200	ည္သ
ပ္ပ	99
FGF	ACA
<b>PGC</b> 6	ည်
ŢŢ	AAG
AAG	FF
25	GTG
999	ညည
AA (	LL
GTG	200
GCGAC	CGCI
71	

Treature		TCACCEAC	LAGTGGATG
COCTOCACIT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTCGACTC		141 CCTGAAGTIC ATCIGCACCA CCGGCAAGUI GCCCGIGCCC IGGCCCACCC ICGIGAACAA CIICACIAC	pasymynssa maancamaan aannamyness neaganandaa senadagamaa senadagama
GCTCCCCCTA		Teecce Acce.	ACCEGETEEG
Geccecarcc		CCCGRECCC	CGGGCACGGG
AAGTCGCACA		CCGCCAAGCT	でいた。日本でいってい
GCCGGTGTTC	- 1	ATCTGCACCA	中にはいるのでは、
CGCTGCACTT		CCTCAAGTIC	じるないたましなりで
	,	141	

フザインフザフィナン	GGACTYCAAG TAGACGTGGT GGCCGTTCGA CGGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG	じじてじて
TCGTGTCGT	AGCACTGGTG	
TGGCCCALCC	ACCGGGTGGG	
つつしていている	CGGGCACGGG	
こうりませつううしつ	GGCCGTTCGA	
ATCIGCACCA	TAGACGTGGT	
41 CCMBARGING AICHGCACCA CCGGCAAGGI GCCCGIGCCC IGGCCCCACCC ICGIGACCAC CIICACCIAC	GGACTTCAAG	1
41		,

A P.	CCGAGGTGAA GGCTCCACTI	AAGACCCGCG TTCTGGGCGC	CGGCAACTAC GCCGTTGATG	TCAAGGACGA AGTTCCTGCT	accatcttct tggtagaaga	AAGGCTACGT CCAGGAGGG ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA TTCCGATGCA GGTCCTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT	281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA TTCCGATGCA GGTCCTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCTTT	281
wh. = 1	GCCATGCCCG	Cttcaagtcc Gaagttcagg	agcaccactit Tcgtgctgaa	CACATGAAGC GTGTACTTCG	CTACCCCGAC GATGGGGCTG	GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG CCGCACGTCA CGAAGTCGGC GATGGGGGTG GTGTACTTCG TCGTGCTGAA GAAGTTCAGG CGGTACGGGC	211 GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA GAAGTTCAGG CGGTACGGGC	211

AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA TTCCGATGCA GGTCCTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT	CGGCAACATC
AAGACCCGCG TTCTGGGCGC	TCAAGGAGGA AGTTCCTCCT
CGGCAACTAC GCCGTTGATG	GGCATCGACT CCGTAGCTGA
TCAAGGACGA AGTTCCTGCT	CGAGCTGAAG
ACCATCTTCT TGGTAGAAGA	TGAACCGCAT
CCAGGAGCGC	GACACCCTGG
281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA TTCCGATGCA GGTCCTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT	351 GTTCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC CAAGGTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG
281	351

CGCCAACATC	Cagaagaacg
GCCGTTGTAG	Gtcttcttgc
GITCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC	CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG
CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG	GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTC GTCTTCTTGC
GGCATCGACT	TCTATATCAT
CCGTAGCTGA	AGATATAGTA
CGAGCTGAAG	AGCCACAACO
GCTCGACTTC	TCGGTGTTGC
TGAACCGCAT	CAACTACAAC
ACTTGGCGTA	GTTGATGTTG
GACACCCTGG	AGCTGGAGTA TCGACCTCAT
351 GITCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCCTCCT GCCGTTGTAG	421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG GACCCCGTGT TCGACCTTGT TCGATGTTG TCGATGTTGC AGATATAGTA CCGGCTGTTC GTCTTTGC
351	421

	ACCACTACCA	CGTAGTICCA CTIGAAGTIC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT
	CAGCTCGCCG	GICGAGCGGC
	CGGCAGCGTG	GCCGTCGCAC
	ACATCGAGGA	TGTAGCTCCT
	491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG CAGCTCGCCG ACCACTACCA	TAGGCGGTGT
-	GAACTTCAAG	CTTGAAGTTC
	GCATCAAGGT	CGTAGTTCCA
	491	

CCGGGATCA	GCCCTAGT
GTGACCGCC G	こうしょうしゅうしょう
TGCTGGAGTT (	ACGACCTCAA C
CACATGGTCC	GTGTACCAGG
GAAGCGCGAT	CTTCGCGCTA
ACCCCAACGA	TCCCCTTCCT
631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA	CACTORATION TO THE CONTROL OF THE GROUP OF TAKES OF A COACTORA COACTIGOR COCCONAGE
631	

# -1G.\_41B

Smal

GAAACCACTG AAGGATGAGC CTTTGGTGAC TTCCTACTCG CGCCGGCGGG CCCGACGTCC CCTGCTCGAC ATGTTCATTT GGACGAGCTG SAGTGCCGTA CTCACGGCAT 701

GTTATTTCAA AGAATTCTAA CTTAGGACAA CGGCCAGAAC GCCGGTCTTG GCATGACGTT GAATCCTGTT AACATGTAAT TGTAATAATT TCTTAAGATT CAATAAAGTT ACGTTAAGCA CTGTTGAATT AAACATTTGG GTCTAGCAAG TTTGTAAACC CAGATCGTTC CATATAATTT TGTAAAGAAG ACATTTCTTC CGATGATTAT

771

sctactaata gtatattaaa gacaacttaa tgcaattcgt acattaataa ttgtacatta cgtactgcaa TACATTTAAT ACGCGATAGA AAACAAATA 841

TLTGILLIAT TAAATACTCT ACCCAAAAAT ACTAATCTCA GGGCGTTAAT ATGTAAATTA TGCGCTATCT CCCGCAATTA TGATTAGAGT TGGGTTTTTA ATTTATGAGA 911

BSSHII CIAI HIAGIII

AGATCGATAA GCTTCTAGAG CGAAGATCTC TCTAGCTATT CTATGTTACT GATACAATGA ATTATCGCGC GCGGTGTCAT TGATCCTATT TAATAGCGCG CGCCACAGTA ACTAGGATAA TAGCGCGCAA ATCGCGCGTT 981

Beshil

CGTTTTACAA GCAAAATGTT CACTGGCCGT GTGACCGGCA TACGCGCGCT GCGGGATAIC ACTCAGCATA ATGCGCGCGA TGAGTCGTAT CGCCCTATAG AGCTCCAATT TCGAGGTTAA GCCGGCCACC CGGCCGGTGG 1051

TTCGCCAGCT AAGCGGTCGA TGTAGGGGGA ACATCCCCCT CGGAACGTCG GCCTTGCAGC CAACTTAATC (GITGAATTAG ( TGGCGTTACC ( CCCTTTTGGG GGGAAAACCC CGTCGTGACT GCAGCACTGA 1121

GCGAATGGGA CGCTTACCCT AGCCTGAATG TCGGACTTAC ACAGTTGCGC TGTCAACGCG GCCCTTCCCA CGGGAAGGGT GCGTGGCTAG CGCACCGATC CGAAGAGGCC GCTTCTCGG GGCGTAATAG CCGCATTATC 1191

ATGTGAACGG TACACTTGCC CGCACTGGCG GCGTGACCGC GTTACGCGCA CAATGCGCGT GGCTGTGGTG TCGCCGCGTA ATTCGCGCCG CCCACACCAC TAAGCGCGGC AGCGGCGCAT CGCGCCCTGT GCGCGGGACA 1261

# -1G.\_41C

1331	AGCGCCCTAG TCGCGGGATC	CGCCCGCTCC	TTTCGCTTTC AAAGCGAAAG	TTCCCTTCCT AAGGGAAGGA	TTCTCGCCAC AAGAGCGGTG	OTTCGCCGGC CAAGCGGCCG	titccccgtc aaaggggcag
1401	aagctctaaa Ttcgagattt	TCGGGGGCTC	CCTTTAGGGT	TCCGATTTAG AGGCTAAATC	tgctttacgg acgaaatgcc	CACCTCGACC GTGGAGCTGG	CCAAAAACT GGTTTTTGA
1471	tgattagggt actaatccca	GATGGTTCAC CTACCAAGTG	GTAGTGGGCC	ATCGCCCTGA TAGCGGGACT	TAGACGGTTT ATCTGCCAAA	ttcgcccttt Aagcgggaaa	gacgttggag Ctgcaacctc
1541	TCCACGTTCT AGGTGCAAGA	ttaatagtgg aattatcacc	ACTCTTGTTC TGAGAACAAG	Caaactggaa Gtttgacctt	CAACACTCAA GTTGTGAGTT	CCCTATCTCG GGGATAGAGC	GTCTATTCTT CAGATAAGAA
1611	ttgalttata Aactaaatat	agggattttg TCCCTAAAAC	CCGATTTCGG GGCTAAAGCC	CCTATTGGTT GGATAACCAA	aaaaaatgag Ttttttactc	CTGATTTAAC GACTAAATTG	aaaaatutaa Tututaaait
1681	CGCGAATTTT GCGCTTAAAA	aacaaaatat Ttgititiata	Taacgettac Attgegaatg	aatttaggtg ttaaatccac	GCACTTTTCG GGGAAATGTG CGTGAAAAGC CCCTTTACAC	GGGAAATGTG CCCTTTACAC	CGCGGAACCC GCGCCTTGGG
1751	Ctatttgitt Gataaacaaa	atititictaa Taaaaagatit	ATACATTCAA TATGTAAGTT	atatgtatcc Tatacatagg	GCTCATGAGA	Caataaccet Getaetegga	gataaatget Ctatttaega
1821	TCAATAATAT AGTTATTATA	Tgaaaaagga actittttcct	AGAGTATGAG TCTCATACTC	tattcaacat Ataagttgta	TTCCGTGTCG AAGGCACAGC	CCCTTATTCC	Cttttttgcg Gaaraacgc
1891	GCATTTTGCC CGTAAAACGG	TTCCTGTTTT AAGGACAAAA	TGCTCACCCA ACGAGTGGGT	GAAACGCTGG CTTTGCGACC	Tgaragtara Actitecatet	agatgctgaa Tctacgactt	GATCAGTTGG CTAGTCAACC
1961	GTGCACGAGT CACGTGCTCA	GGGTTACATC	GAACTGGATC CTTGACCTAG	TCAACAGCGG AGTTGTCGCC	TAAGATCCTT ATTCTAGGAA	gagagtititc Ctctcaaaag	GCCCCGAAGA CGGGGCTTCT
2031	acgitticca tgcaaaaggt	ATGATGAGCA TACTACTCGT	Ctttttaaagt Gaaaatttca	TCTGCTATGT AGACGATACA	GGCGCGGTAT	Tatcccgtat Atagggcata	TGACGCCGGG
2101	CAAGAGCAAC	CAAGAGCAAC TCGGTCGCCG GTTCTCGTTG AGCCAGCGGC	CATACACTAT GTATGTGATA	TCTCAGAATG AGAGTCTTAC		ACTTGGTTGA GTACTCACCA TGAACCAACT CATGAGTGGT	GTCACAGAAA CAGTGTCTTT

# FIG.\_41D

2171	AGCATCTTAC TCGTAGAATG	GGATGGCATG CCTACCGTAC				accatgagtg Tggtactcac	ataacactgc tattgtgacg
2241	GGCCAACTTA	CTTCTGACAA	CGATCGGAGG	acceaageag	CTAACCGCTT	TTTTGCACAA	Categegeat
	CCGGTTGAAT	GAAGACTGTT	GCTAGCCTCC	Tggcttcctc	GATTGGCGAA	AAAACGTGTT	Gtaccccta
2311	Catgtaactc	GCCTTGATCG	TTGGGAACCG	gagctgaatg	AAGCCATACC	AAACGACGAG	CGTGACACCA
	Gtacattgag	CGGAACTAGC	AACCCTTGGC	ctcgacttac	TTCGGTATGG	TTTGCTGCTC	GCACTGTGGT
2381	CGATGCCTGT	agcaatggca	ACAACGTTGC	gcaaactatt	AACTGGCGAA	Ctacttactc	TAGCTTCCC9
	GCTACGGACA	Tcgttaccgt	TGTTGCAACG	cgtttgataa	TTGACCGCTT	Gatgaatgag	ATCGAAGGGC
2451	GCAACAATTA CGTTGTTAAT	atagactega tatctgacct	TGGAGGCGGA	taaagttgca atttcaacgt	GGACCACTIC CCTGGTGAAG	TGCGCTCGGC ACGCGAGCCG	CCTTCCGCT
2521	GGCTGGTTTA CCGACCAAAT	ttgctgataa Aacgactatt	atctggagcc tagacctcgg	GGTGAGCGTG CCACTCGCAC	GGTCTCGCGG CCAGAGCGCC	Tatcattgca Atagtaacgt	GCACTGGGGC
2591	Cagatggtaa Gtctaccatt	GCCCTCCCGT	atcgtagtta tagcatcaat	TCTACACGAC AGATGTGCTG	GGGGAGTCAG CCCCTCAGTC	gcaactatgg cgttgatacc	atgaacgaaa tacttgcttt
2661	TAGACAGATC	GCTGAGATAG	GTGCCTCACT	gattaagcat	TGGTAACTGT	CAGACCAAGT	ttactcatat
	ATCTGTCTAG	CGACTCTATC	CACGGAGTGA	Ctaattcgta	ACCATTGACA	GTCTGGTTCA	Aatgagtata
2731	atactttaga	ttgatttaaa	acttcatttt	taatttaaaa	GGATCTAGGT	Gaagatcctt	tttgataatc
	tatgaaatct	Aactaaattt	Tgaagtaaaa	Attaaaittt	CCTAGATCCA	Cttctaggaa	aaactattag
2801	TCATGACCAA	aatcccttaa	CGTGAGTTTT	CGTTCCACTG	AGCGTCAGAC	CCCGTAGAAA	agatcaaagg
	AGTACTGGTT	Ttagggaatt	GCACTCAAAA	GCAAGGTGAC	TCGCAGTCTG	GGGCATCTTT	Tctagtttcc
2871	atcttcttga	gatectititi	ticigegege	aatctgctgc	ttgcaaacaa	AAAAACCACC	GCTACCAGCG
	Tagaagaact	Ctaggaaaaa	Aagaegegea	Ttagacgacg	aacgtttgtt	TTTTTGGTGG	CGATGGTCGC
2941	GTGGTTTGTT	TGCCGGATCA	AGAGCTACCA	actettttttc	CGAAGGTAAC	TGGCTTCAGC	agagegerga
	CACCAAACAA	ACGGCCTAGT	TCTCGATGGT	Tgagaaaaag	GCTTCCATTG	ACCGAAGTCG	Tetegegtet

#### FIG.\_41E

1011 TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGCCGT GCTGAAGGTC TACCGACTAG CACCGCCTAC ATGGTTATAG ACACGAAATAC TGTCCTATAC ACAGGAAGAT CACAGGAAGAT CACAGGAAGTC GTGGCGGATG TACGGGTTG TACGGGTTG TACGGGTTG TACGGGTTG TACGGGTTG TACGGGTTG TACGGGTTG TACGGGTTG TACGGGTTG TATGGAACTC CTGTTACCATG GCTGAACGGG GCTGAACGGG GGTTCGTCG TACGGGTTG TACGGGTTG TACGGGTTG CTGTCTAGACG ACACAGCCCAAC GAGGGTTG TACGGGTTG TACGTTGTTC TACGGGTTG TACGTTGTTC TACGGGTTG TACGGTTG TACGTTGTTC TACGGGTTG TACGTTGTTC TACGGTTGT TA	A TG	AC A	ឡុខ	ပ္ပ ဗ္ဗ	TT AA	TT AA	ပ္ပ စ္	ပ္ပ စ္တ	P P	CT B	85 E	ජ
TACCAATAC TGTCCTTCTA GTGTAGGCCGT AGTTAGGCCA CCACTTCAAG AACTCGTAGAATGAC ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCGGGT GGTGAAGGG TTGAGACAGAATGAC TATGGGTCGA GTGAACGGG GGGTTCGTGT GTGTAGGACGG GGGTTCGTGT GTGAGAGGG GTTGAGACGG GGGTTCGTGT TATGGAGACGG GTTGATCGTG GTGAGAGGG GGGTTCGTGTG GTGAGAGGG GTTGAGACGG GTGAGAGGG GGGTTCGTGGTGTGTGTGTGTGTGTGTGTGTGTGTG	CACCGCCT	TACCGGGT	ACACAGCC TGTGTCGG	CCACGCTT	gagggagc Ctcctcg	CGTCGAIT	GGTTCCTG	CCGTATTA GGCATAAT	agcgagga Tcgctcct	GCTGGCAC	TCATTAGG AGTAATCC	CAATTTCA
TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTCAAGA ATGGTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC  ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC  ATAGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT  GACTTGGAGCG AACGGACTAG CCTATTCCGC GTGCCCACCC  GCTTGGAGCGA AACGACTAC ACCGAACTGA GATACCTACA GCGTGAACGG  GCTTGGAGCGA AACGACTAC ACCGAACTGA GATACCTACA GCGTGAACGG  GCTTGGAGGAGA AACGCCTGT ACCGAACTGA GATACCTACA GCGTGAACAG  GCTTGCCTCT TTCCGCCTGT TCTTTATAGT CTTGCCGTCC CAGCCTTGTC  GGTACCCCCTT TGCGGACAT AGAAATATCA GGACAGCCCA AACGGTGAA  TGTGAAGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCCACC  GGTCCCCCTT TGCGGACCAT AGAAATATCA GGACAGCCCA AACGGTGAA  TGTGAAGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGGTGAA  TGTGAAGGGGAA ACGCCTGGTA GAAAATATCA GGACACGCCA  GCTTTTGCTGG GTCAGGGGG CGCAGCCTAT GGCGAACGC  CCTTTTGCTGG CAGTCCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT  GCAAAAACGGC CAGTCACAC GCTCGCCGC AGCCGAACGC  CTTTTGCTGA GGCTGATAC GCTCGCCGC AGCCGAACGA  GCCTTTGAGC GGAAAAACGA TTTAGGCGCAACGC AGCCGAACGA  GCCTTTGAGC GGATATGCG TTTGGCGCAACGC AACCGCCTAA  CGGAAAAACGC CCCAATACG GCGCGCAACGC AACCGCCAACGC  GCCTTTTAGCTG GGACCTATA GCCTCCCCCCCC GCCTCGCTTACTCT GCTCGCGTAA  GCCTTTTAGCTG GGACTGATAC GCGACCGCC AACCGCCAACGCC AACCGCCCAACGC AGCCCCAACGC AGCCCTCAACGC AGCCCCAACGC AACCGCCTCAACGC AGCCCCCCCCCC	AACTCTGTAG TTGAGACATC		GGGTTCGTGC CCCAAGCACG	TGAGAAAGCG ACTCTTTCGC	GAGAGCGCAC CTCTCGCGTG	CTGACTTGAG GACTGAACTC	GCCTTTTTAC CGGAAAAATG	CTGTGGATAA GACACCTATT		CATTAATGCA GTAATTACGT	GTTAGCTCAC CAATCGAGTG	GAGCGGATAA
TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCG TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG GACTCCAGGC AACGATTAGG CCTATTCCGC GTCGCCAGCC GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA CGAAGGAGA AACGCCTGT TGCCGCTTCT CCATAGGCCA TTCGCCGTCC CCAGGGGGAA ACGCCTGTA TCTTTATAGT CCTGTCGGTC GCTTCCCCTT TGCGGACCTA AGAATATCA GGAAAAACGC ACATAGGGGAA ACGCCTGGTA TCTTTATAGT CCTGTCGGT GGTCCCCTT TGCGGACCTAT AGAAATATCA GGAAAAACGC ACATACGAG CAGTCCCCC GCCTCGGATA TGTGATGCTGG CAGTCCCCC GCCTCGGATA GGCTTTGAGT GAAAAACAAA GGAAAACGC ACACTACGAG CAGTCGCTAT GCGAAAACGC ACACTACGAG CAGTCGCTAT GCGAAAACGC ACACTACGAG CAGTCGCTAT GCGAAAACGC ACACTACGAG CAGTCGATAC GCCTCGCGTAT GCGAAAACGC CGAAAAACGAG TGTACAAAGA AGGACGCAACG CCGAAAACCCC GGAAAACGC AAACCGCCCC AGCCGCGCG GCCTTTTGAGT GAGTTATGCG TTTGGCGAAG CGGAAACTCC GGGTTATGCG TTTGGCGAAG CCGAAAGGG CTGACTATC GCCCCCCCC CCGAAGGGG CTGACTTTTC GCCCCCCCCCCCCCCCCCC	CCACTTCAAG GGTGAAGTTC	agtggcgata Tcaccgctat	GCTGAACGGG CGACTTGCCC	GCGTGAGCTA CGCACTCGAT	GTCGGAACAG CAGCCTTGTC	TTCGCCACCT AAGCGGTGGA	CAGCAACGCG GTCGTTGCGC	TCCCCTGATT AGGGGACTAA	CCGAGCGCAG GGCTCGCGTC	TTGGCCGATT AACCGGCTAA	attratgega Taaptacact	
TACCAAATAC TGTCCTTCTA GTGTAGCCGT ATGGTTTATG ACAGGAAGAT CACATCGGCA ATGGTTTATG ACAGGAAGAT CACATCGGCG CTGAGTTCTG CTGCTAATCC GGATAAGGCG CTGAGTTCTG CTATCAATGG CCTATTCCGC GCTTGGGAGA AACGCGGACA GGTATCCGGT GCTTCCCTCT TTCCGCCTGT CCTTTATAGT GCTTCCCTCT TTCCGCCTGT CCTTTATAGT GCTTCCCCTT TGCGGACCTT TCTTTATAGT GCTTCCCCTT TGCGGAGGGG CGGAGCCTAT ACACTACGAG CAGTCCCCC GCTTTTGCTG CTTTTGCTC GCTTTTGCTG CCTTTTGCT GGAAATTCT GGGAAACGAC CGGAAACGAC CGGAAACGAC CGGAAACGC CGGAAGAGC CCCTTTTGCC GCCTTTTGCT GGGAAACGC CGGAAACGC CGGAAACGC CCCAATACGC CGGAAGAGC CCCAATACGC CGGAAGAGC CCCCAATACGC CCCCAATACGC CCCCAATACGC CCCCAATACGC CCCCCCCCC CCCCCCCCC CCCCCCCCCC	agttaggcca Tcaatccggt		CAGCGGTCGG GTCGCCAGCC	gatacctaca Ctatggatgt			ggaaaaacgc cctttttgcg	TCCTGCGTTA AGGACGCAAT			GCGCAACGCA CGCGTTGCGT	CGTATGTTGT
TACCAAATAC TGTCCTTCTA ATGGTTTATG ACAGGAAGAT ATTGGAGCGA GACGATTAGG GACTCAAGAC GATAGTTAGG GACTCAAGAC GATAGTTAGG GCTTGGAGGAGA AAGGCGGACA GCTTCCCTCT TTCCGCCTGTA GCTTCCCTCT TTCCGCCTGTA GCTTCCCTCT TTCGGCCTGTA GCTTCCCTTT TGCGCACTA CCAGGGGGAA ACGCCTGGTA GCTTCCCTT TGCGCACTAT CCAGGGGGAA ACGCCTGTC CTTTTGCTC GTCAGGGGGGG ACACTCCCCT CTTTTGCTCG CTCTTTGCTC GAAAACGAG CCCTTTTGCTC GAAAACGAG CCCTTTTGCTC GAAAACGAG CCCTTTTGCTC GCGAAGAGCG CCCTATTGCTC CCGGAAGAGCG CCCAATACGC GCCTTTTCCC GACTGGAAAG CCGCAATACCC CCCCAATACGC GCCTTTTCCC GACTGGAAAG CCCCCAATACGC GCCTTCTCCC GACTGGAAAG CCCCCAATACGC CCCCAATACGC CCCCAATACCC CCCCCCCTTTTTCCCCC CCCCCCTTTTTCCCCCCC												
				AACGACCTAC TTGCTGGATG				CCTTTTGCTC GGAAAACGAG	gagctgatac Ctcgactatg	CCCAATACGC GGGTTATGCG		
	TACCAAATAC ATGGTTTATG										CAGGTTTCCC	CCCCAGGCTT
	3011	3081	3151	3221		3361	_ •					

-1G.\_41F

3851	CAGGAAACAG GTCCTTTGTC	CTATGACCAT GATACTGGTA	Gattacgcca Ctaatgcggt	Babhii Agccccaat Tccccctra	TAACCCTCAC Attgggagtg	Taaagggaac atttcccttg	ECORI ~~ AAAAGCTGGA T'TTTCGACCT
3921	ECORI ~~~ ATTCCACAAT TAAGGTGTTA	Gaacaataat Cttgttatta	aagattaaaa Ttctaatttt	TAGCTTGCCC ATCGAACGGG	CCGTTGCAGC	gatgggtatt Ctacccataa	tittctagta Aaaagatcat
3991	aaataaaga Titatitict	taaacttaga Atttgaatct	CTCAAAACAT GAGTTTTGTA	ttacaaaac aatgittitg	AACCCCTAAA TTGGGGATTT	GTCCTAAAGC CAGGATTTCG	CCAAAGTGCT GGTTTCACGA
4061	ATGCACGATC TACGTGCTAG	Catagcaagc Gtatcgttcg	CCAGCCCAAC	CCAACCCAAC GGTTGGGTTG	CCAACCCACC GGTTGGGTGG	CCAGTGCAGC GGTCACGTCG	Caactggcaa Gttgaccgtt
4131	ATAGECECCA TATCAGAGGE	CCCCGGCAC	TATCACCGTG ATAGTGGCAC	AGTTGTCCGC TCAACAGGCG	ACCACCGCAC TGGTGGCGTG	GTCTCGCAGC	Caaaaaaaa Giititititi
4201	aaaagaaaga Titicitict		aaagaaaaac Tttctttttg		TCCGGGTCGT	GGGGGCCGGA	aaagcgagga Titcgcicci
4271	GGATCGCGAG CCTAGCGCTC		CAGCGACGAG GCCCGGCCT CCCTCCGCTT GTCGCTGCTC CGGGCCGGGA GGGAGGCGAA		CCAAAGAAAC GGTTTCTTTG	GCCCCCCATC	gccactatat Cggtgatata
4341	ACATACCCCC TGTATGGGGG	CCCTCTCTCTC	CCATCCCCC AACCCTACCA GGTAGGGGGG TTGGGATGGT		CCACCACCAC	CACCACCTCC GTGGTGGAGG	TCCCCCTCG
4411	CTGCCGGACG	ACGAGCTCCT TGCTCGAGGA	CCCCCTCCC	CCTCCGCCGC	CGCCGGTAAC GCGGCCATTG	CACCCGCCC	CTCTCCTCTT GAGGAGAA
4481	TCTTTCTCCG AGAAGAGGC	tttttttt Aaaaaaaaa	CGTCTCGGTC	TCGATCTTTG AGCTAGAAAC	GCCTTGGTAG	TTTGGGTGGG	CGAGAGCGGC GCTCTCGCCG
4551	TTCGTCGCCC AAGCAGCGGG	AGATCGGTGC TCTAGCCACG	GCCCTCCC	CGGGATCTCG CGGCTGGCGT GCCCTAGAGC GCCGACCGCA		CTCCGGGCGT	Grancogece Ctcrocegg

FIG.\_41G

	*****			5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5			
4621		GGGGAATGGG CCCCTTACCC	GGATCCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTTC TTTCTTTCTT CTTTTTGG TAGAATTTGA CCTAGGAGCG CCCCTTACCC CGAGAGCCTA CATCTAGAAG AAAGAAAGAA GAAAAACACC ATCTTAAACT	gtagatcttc Catctagaag	titciticti Aaagaaagaa	Cttttttgtgg Gaaaaacacc	Tagaatitiga Atcitaaact
4691	ATCCCTCAGC TAGGGAGTCG	ATTGTTCATC TAACAAGTAG	ATCCCTCAGC ATTGITCATC GGIAGTITIT CITTICATGA TITGIGACAA ATGCAGCCTC GIGCGGAGCT TAGGGAGTCG TAACAAGTAG CCATCAAAA GAAAAGTACT AAACACTGIT TACGICGGAG CACGCCTCGA	CTTTTCATGA GAAAAGTACT	tttgtgacaa Aaacactgtt	atgcagcctc tacgtcggag	GTGCGGAGCT
4761	4761 TTTTTGTAGG TAG	TAG	-				

BamHI

FIG.\_41H

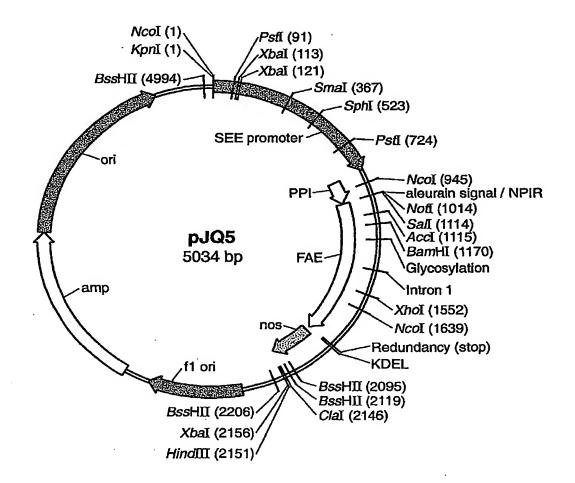


FIG.\_42A

ਜ	Catgggccag Gtacccggtc	GTATAATTAT CATATTAATA	CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG GTACCCGGTC CATATTAATA CCCTATAGAG TTCGTTTATT AGCTTTATAG TGGTAACCGA TGTTATAGAC	aagcaaataa TTCGTTTATT	TCGAAATATC AGCTTTATAG	ACCATTGGCT TGGTAACCGA	ACAATATCTG TGTTATAGAC
		Pati	1		XbaI	XbaI	
71	agctccgagt Tcgaggctca	TCTGACTGCA AGACTGACGT	AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTTGT TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAACA	CGCGTGTTGT GCGCACAACA	ATCTAGAACT TAGATCTTGA	ATCTAGAACT CTAGATAGCA CAGCCACAGG TAGATCTTGA GATCTATCGT GTCGGTGTCG	CAGCCACAGC GTCGGTGTCG
141	acctacagga Tggatgtcct	GTGCGACACT CACGCTGTGA	ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTTG GAGACGGAGC TCTTTCCTAC TGGATGTCCT CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCTCG AGAAAGGATG	agtagtgttg Tcatcacaac	GAGACGGAGC CTCTGCCTCG	TCTTTCCTAC AGAAAGGATG	CTCCTGACGT GAGGACTGCA .
211	TGCCGCCGTT ACGGCGGCAA	GTCCATTCCA CAGGTAAGGT	TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC ACGGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTGCGCGAG GGTTGTTTTA TAGCAGGGG	TCTCAACCAA AGAGTTGGTT	TCACGCGCTC AGTGCGCGAG	CCAACAAAT GGTTGTTTTA	atcetcccc tagcaggggg
281	ATGTCTTGGC TACAGAACCG	GGAGAGAGAG	ATGTCTTGGC GGAGAGAGA TACATACATG CTGTCGCGCC GTTTTTGTCT GAATCTCGCT TCCACTGGCC TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAAACAGA CTTAGAGCGA AGGTGACCGG	CTGTCGCGCC GACAGCGCGG	gitititgect Caaaaacaga	GITITITGICT GAATCICGCT TCCACTGGCC CAAAAACAGA CTTAGAGCGA AGGIGACCGG	TCCACTGGCC AGGTGACCGG
351	aatcagctca Ttagtcgagt	Smal GCTCCGGGA CGAGGGCCCT	Smai ~~~~~ AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCTAGGG TAGCAGCAGC AGTGGGGAACC	TCAAGATCCC AGTTCTAGGG	ATCGTCGTCG TAGCAGCAGC	TCACCCTGG CGTCATGGGA AGTGGGGAACC GCAGTACCCT	CGTCATGGGA GCAGTACCCT
421	TGGAAAGAA ACCTTTTCTT	CCTCCGTTGC	TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT ACCTTTTCTT GGAGGCAACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTC TATTGGGTTA	CAGCCATATC OTCGGTATAG	CCCGAACAGA GGGCTTGTCT	GTACTGCAAG ATAACCCAAT CATGACGTTC TATTGGGTTA	ataacccaat Tattgggtta
491	TCAGATTCCC	CCAATAGAGA GGTTATCTCT	Spbi TCAGATTCCC CCAATAGAA AAGTATAGCA TGCTTTCGGG TTTTGTTTGG CTTAATTGAC TTTATTTTTG AGTCTAAGGG GGTTATCTCT TTCATATCGT ACGAAAGCCC AAAACAAACC GAATTAACTG AAATAAAAAC	II TGCTTTCGGG ACGAAGCCC	ttttgtttgg Aaaacaaacc	CTTAATTGAC GAATTAACTG	tttattttg Aaataaaaac

FIG.\_ 42

TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG AACCTCAACT TACGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCTA TTATCCGACC

561

631	Ctaattaatt Gattaattaa	Tatagcaaga Atatcgttct	CTAATTAATT TATAGCAAGA TYCTGTAGTG CACATGGCAA GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT	CACATCGCAA GTGTAGCGTT		ATATCTTTCT GGGCATTACA TATAGAAAGA CCCGTAATGT	GCTGGAGGCT CGACCTCCGA
		Ž į	10 10 10 10 10 10 10 10 10 10 10 10 10 1				
701	TCATCAGCCT AGTAGTCGGA		GAAACACTCT GCAGAGCCTG CTTTGTGAGA CGTCTCGGAC	AAGCAAGTGG TTCGTTCACC	TGAAGCGTGG ACTTCGCACC	Cgatgagatg Gctactctac	GGTATAAAAC CCATATTTTG
771	CCCCGGCCACC	GGGACGCGAG	CTCCCGCCTA		CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGTCATGGTA QAGCGGAGCG AGGGGGACGG	TCCCCCTGCC	GGACGACCCA CCTGCTGGGT
841	gtaaaatact Catititatga	GTTGCCCACT CAACGGGTGA			ATGGMCGTGC ACAAGGAGGT TACCKGCACG TGTTCCTCCA	Saacticges Stigaagcas	GCCTACCTCC CGGATGGAGG
911	TGATCGTSCT ACTAGCASGA	CGGCCTCCTC	ttgctcgtst aacgagcasa	Ncol CCGCCATGGA GGCGGTACCT	GCACGTGGAC CGTGCACCTG	GCCAAGGCCT GCACCCKCGA CGGTTCCGGA CGTGGGMGCT	GCACCCKCGA
981	GTGCGGCAAC	CTCGGCTTCG	GCATCTGCCC	Not I Gececcecc	Noti GGCGGCCCC TCCACGCAGG GCATCTCCGA CCGCCGGCGG AGGTGCGTCC CGTAGAGGCT	GCATCTCCGA CGTAGAGGCT	agacctctac Tctggagatg
							Sali
1051	agccotttag Tcgccaaatc	TCGAAATGGC AGCTTTACCG		CACTATCTCC CAAGCTGCCT ACGCCGACCT GTGCAACATT GTGATAGAGG GTTCGACGGA TGCGGCTGGA CACGTTGTAA	CAAGCTGCCT ACGCCGACCT OTTCGACGGA TGCGGCTGGA	GTGCAACATT CACGTTGTAA	CCGTCGACTA GGCAGCTGAT
1121	TTATCAAGGG AATAGTTCCC	TTATCAAGG AGAGAAAATT AATAGTTCCC TCTCTTTTAA	tacaattctc atgttaagag		Taacggat	Bamei GG ATCCTCCGCG CC TAGGAGGCGC	acgacagcag tgctgtcgtc

1191	Caragarata Giticitiri		ATCACCETCT TCCGTGGCAC TGGTAGTGAT TAGTGGCAGA AGGCACCGTG ACCATCACTA	tggtagtgat accatcacta	ACGAATCTAC TGCTTAGATG	AACTCGATAC TTGAGCTATG	Taactacacc Attgatgtgg
1261	CTCACGCCTT GAGTGCGGAA	TCGACACCCT AGCTGTGGGA	accacaatgc tggtgttacg	AACGGTTGTG TTGCCAACAC	AAGTACACGG TTCATGTGCC	tggatattat Acctataata	attggatggg taacctaccc
1331	TCTCCGTCCA AGAGGCAGGT	GGACCAAGTC CCTGGTTCAG	GAGTCGCTTG TCAAACAGCA CTCAGCGAAC AGTTTGTCGT	TCAAACAGCA AGTTTGTCGT	GGTTAGCCAG CCAATCGGTC	TATCCGGACT ATAGGCCTGA	ACGCGCTGAC TGCGCGACTG
1401	CGTGACCGGC	CACKCCCTCG GTGMGGGAGC	GCGCCTCCCT GGCGGCACTC CGCGGAGGGA CCGCCGTGAG		ACTGCCGCCC AGCTGTCTGC TGACGGCGGG TCGACAGACG	AGCTGTCTGC TCGACAGACG	Gacatacgac Ctgtatgctg
1471	AACATCCGCC TTGTAGGCGG	TGTACACCTT ACATGTGGAA	CGGCGAACCG CGCAGCGGCA GCCGCTTGGC GCGTCGCCGT	CGCAGCGGCA	atcaggcctt Tagtccggaa	CGCGTCGTAC GCGCAGCATG	atgaacgatg tacttgctac
17 17 17	CCTTCCAAGC GGAAGGTTCG	XhoI CTCGAGCCCA GAGCTCGGGT	GATACGACGC CTATGCTGCG	agtatttccg TCATAAAGGC	GGTCACTCAT GCCAACGACG CCAGTGAGTA CGGTTGCTGC	GCCAACGACG CGGTTGCTGC	GCATCCCAAA CGTAGGGTTT
1611	CCTGCCCCG	CCTGCCCCC GTGGAGCAGG	Nec	ol TGGCGGTGTA	GAGTACTGGA	GCGTTGATCC	TRACAGOGOC
1681	CAGAACACAT		TGGGGATGAA	GTGCAGTGCT	GTGAGGCCCA CACTCCGGGT	GGGGGGACAG	GGTGTGAATA CCACACTTAT
1751	ATGCGCACAC TACGCGTGTG	gacttattit Ctgaataaaa	GGGATGACGA GCGGAGCCTG CCCTACTGCT CGCCTCGGAC		TACATGGTGA TCAGTCATTT ATGTACCACT AGTCAGTAAA	TCAGTCATT AGTCAGTAAA	CAGCCTCCCC
1821	GAGTGTACCA CTCACATGGT	ggaaagatgg Cctttctacc	ATGTCCTGGA TACAGGACCT	GAGGGGGCCG CTCCCCCGGC	CGTAACCACT GAAGGATGAG GCATTGGTGA CTTCCTACTC	gaaggatgag Cttcctactc	Ctgtaaagaa Gacatttctt

### FIG.\_42D

ATTA FAAT	TGAG	HII	GCAC	TGAC	aata Itat	CCTG	CCTA
gcgatgatta Cgctactaat	tatttatgag Ataaatactc	BBBHII ~~~~~ ATAGCGCGCA TATCGCGCGT	GCGGCCAC	ACGTCGTGAC TGCAGCACTG	TGGCGTAATA ACCGCATTAT	<b>ACGCGCCCTG</b> TGCGCGGGAC	CAGCGCCCTA GTCGCGGGAT
CAAACATTIG GCAATAAAGT TICTTAAGAT TGAATCCIGT TGCCGGTCTT GITTGTAAAC CGTTATTICA AAGAATTCTA ACTTAGGACA ACGGCCAGAA	ATGTAATAAT TAACATGTAA TGCATGACGT TACATTATTA ATTGTACATT ACGTACTGCA	ATGATTAGAG TCCCGCAATT ATACATTTAA TACGCGATAG AAAACAAAAT TACTAATCTC AGGGCGTTAA TATGTAAATT ATGCGCTATC TTTTGTTTTA	Xbal Clal Hindili TAGATCGATA AGCTTCTAGA	TCACTGGCCG TCGTTTTACA ACGTCGTGAC AGTGACCGGC AGCAAAATGT TGCAGCACTG	TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG CACATCCCCC TTTCGCCAGC TGGCGTAATA ACCCTTTTGG GACCGCAATG GGTTGAATTA GCGGAACGTC GTGTAGGGGG AAAGCGGTCG ACCGCATTAT	CCGCACCGAT CGCCCTTCCC AACAGTTGCG CAGCCTGAAT GGCGAATGGG ACGCGCCCTG GGCGTGGCTA GCGGGAAGGG TTGTCAACGC GTCGGACTTA CCGCTTACCC TGCGCGGGAC	TAGCGGCGCA TTAAGCGCGG CGGGTGTGGT GGTTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCTA ATCGCCGCGT AATTCGCGCC GCCCACACCA CCAATGCGCG TCGCACTGGC GATGTGAACG GTCGCGGGAT
TGAATCCTGT ACTTAGGACA	Taacatgtaa Attgtacatt	TACGCGATAG	Clar Bindiri TAGATCGATA AGCTT ATCTAGCTAT TCGAA		CACATCCCCC GTGTAGGGGG	CAGCCTGAAT GTCGGACTTA	AGCGTGACCG
ttcttaagat Aagaattcta	TACGTTAAGC ATGTAATAAT ATGCAATTCG TACATTATTA	atacatttaa Tatgtaaatt	TCTATGTTAC AGATACAATG	BSSHII TTACGCGCGC AATGCGCGCG	CGCCTTGCAG GCGGAACGTC	CGCCCTTCCC AACAGTTGCG GCGGGAAGGG TTGTCAACGC	GGTTACGCGC
gcaataaagt Cgttatttca	TCTGTTGAAT TACGTTAAGC AGACAACTTA ATGCAATTCG	TCCCGCAATT AGGGCGTTAA	BSSEII	GTGAGTCGTA CACTCAGCAT	CCAACTTAAT GGTTGAATTA	CGCCCTTCCC	CGGGTGTGGT
Caaacatttg Gittgiaaac		atgattagag tactaatctc	BBBEII  AATTATCGCG CGC TTAATAGCGC GCG	TCGCCCTATA AGCGGGATÄT	CTGGCGTTAC	CCGCACCGAT	TTAAGCGCGG
GCAGATCGTT CGTCTAGCAA	TCATATATT AGTATATTAA	атссстттт Тасссааааа	AACTAGGATA TTGATCCTAT	GAGCTCCAAT CTCGAGGTTA	TGGGAAAACC ACCCTTTTGG	GCGAAGAGGC CGCTTCTCCG	TAGCGGCGCA
1891	1961	2031	2101	2171	2241	2311	2381

### =1G.\_42E

GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT CAAGCTCTAA CGCGGCCAA GAAAGGGGCA GTTCGAGATT

AICGEGEGET CCCTTIAGGE ITCCGAITIA GIECTTIACE GCACCTCGAC CCCAAAAAC TIGAITIAGGG	TGACGTTGGA GTCCACGTTC	GGTCTATTCT TTTGATTTAT	Cararatta acgegratte	GCGCGGAACC CCTATTTGTT	TGATAAATGC TTCAATAATA	CCTTTTTGC GGCATTTTGC	AGATCAGTTG GGTGCACGAG	CGCCCCGAAG AACGTTTTCC	TTGACGCCGG GCAAGAGCAA	agtcacagaa aagcatctta	GATAACACTG CGGCCAACTT
TAGCCCCCCGA GGGAAAICCC AAGGCIAAAI CACGAAAIGC CGIGGAGCIG GGGIITITIIG AACIAAICCC	ACTGCAACCT CAGGTGCAAG	CCAGATAAGA AAACTAAATA	Giftitaari igcgettaar	CGCGCCTTGG GGATAAACAA	ACTATTTACG AAGTTATTAT	GGAAAAACG CCGTAAAACG	TCTAGTCAAC CCACGTGCTC	GCGGGGCTTC TTGCAAAAGG	AACTGCGGCC CGTTCTCGTT	tcagtgtctt ttcgtagaat	
o gercercare co c cercaracie go	TTTCGCCCTT AAAGCGGGAA	ACCCTATCTC TGGGATAGAG	gctgatttaa Cgactaaatt	GGGGAAATGT CCCCTTTACA	ACAATAACCC TGTTATTGGG	GCCCTTATTC CGGGAATAAG	AAGATGCTGA TTCTACGACT	TGAGAGTTTT ACTCTCAAAA		AGTACTCACC TCATGAGTGG	AACCATGAGT
rta Grechtiac Aar Caccaaatg	CTG ATAGACGGTT	GGA ACAACACTCA CCT TGTTGTGAGT	egt taaaaaatga CCA attttttact	GGT GGCACTTTTC CCA CCGTGAAAAG	ATC CGCTCATGAGIAGAGIAG	ACA TTTCCGTGTC IGT AAAGGCACAG	CTG GTGAAAGTAA GAC CACTTTCATT	SCG GTAAGATCCT CGC CATTCTAGGA	TUCTECTATE TEGCECEGTA TTATCCCGTA AAGACGATAC ACCECECAT AATAGGGCAT	AAT GACTIGGIIG ITA CIGAACCAAC	GCA GTGCTGCCAT
CCCTTTRAGGG TICCGATTTR GIGCTTTRACG	CGIAGTGGGC CATCGCCCTG	GACTCTTGTT CCAAACTGGA	GCCGATTICG GCCTATTGGT	ttaacgeita caatitaggi	AATACATTCA YATATGTATC	AAGAGTATGA GTATTCAACA	TTGCTCACCC AGAAACGCTG	CGAACTGGAT CTCAACAGCG	ACTITITAAAG ITCTGCTATG	GCATACACTA TTCTCAGAAT	GACAGTAAGA GAATTATGCA
GGGAAATCCC AAGGCTAAAT CACGAAATGC	GCATCACCCG GTAGCGGGAC	CTGAGAACAA GGTTTGACCT	CGGCTAAAGC CGGATAACCA	aatitgegaat gitaaateca	TTATGTAAGT TTATACATAG	TTCTCATACT CATAAGTTGT	AACGAGTGGG TCTTTGCGAC	GCTTGACCTA GAGTTGTCGC	TGAAAATTTC AAGACGATAC	CGTATGTGAT AAGAGTCTTA	
ATCGGGGGCT CCCTT	TGATGGTTCA CGTAG	ттватасто састс	aagggattti gccga	TAACAAATA TTAAC	татттттста дагас	TTGAAAAGG AAGAG	CTTCCTGTTT TTGCT	TGGGTTACAT CGAAC	AATGATGAGC ACTTI	CTCGGTCGCC GCATA	CGGATGCCAT GACAGTAAGA
TAGCCCCCGA GGGAA	ACTACCAAGT GCATC	Адаттатсас столо	Ticcciaaaa cggci	ATTGTTTAT AATTG	Атаааааат ттат	AACTITTTCC TTCTC	GAAGGACAAA AACGA	ACCCAATGTA GCTTG	TTACTACTCG TGAAA	GAGCCAGCGG CGTAI	
2521 AT	2591 TC AC	2661 TY	2731 AA	2801 TA	2871 TZ	2941 TH	3011 CF	3081 TG	3151 A	3221 CT	3291 CC

96/154

3361	ACTTCTGACA TGAAGACTGT	acgatcggag Tgctagcctc	GACCGAAGGA CTGGCTTCCT	GCTAACCGCT TTTTTGCACA CGATTGGCGA AAAAACGTGT	TTTTGCACA AAAAACGTGT	acategegga Tgtacccct	TCATGTAACT AGTACATTGA
3431	CGCCTTGATC	GTTGGGAACC	ggagctgaat Cctcgactta	GAAGCCATAC CTTCGGTATG	Caracgacga Gitigcigct	GCGTGACACC CGCACTGTGG	acgatgcctg Tgctacggac
3501	TAGCAATGGC ATCGTTACCG	aacaacgttg Ttgttgcaac	CGCAAACTAT GCGTTTGATA	Taactggcga Attgaccgct	actacttact Tgatgaatga	CTAGCTTCCC	ggcaacaatt ccgttgttaa
3571	aatagactgg Ttatctgacc	ATGGAGGCGG TACCTCCGCC	ATAAAGTTGC TATTTCAACG	AGGACCACTT TCCTGGTGAA	CTGCGCTCGG	CCCTTCCGGC GGGAAGGCCG	TGGCTGGTTT ACCGACCAAA
3641	attgctgata taacgactat	aatctggagc ttagacctcg	CGGTGAGCGT GCCACTCGCA	GGGTCTCGCG	gtatcattgc Catagtaacg	AGCACTGGGG TCGTGACCCC	CCAGATGGTA GGTCTACCAT
3711	AGCCCTCCCG TCGGGAGGGC	TATCGTAGTT	ATCTACACGA TAGATGTGCT	CGGGGAGTCA GCCCCTCAGT	GGCAACTATG CCGTTGATAC	gatgaacgaa Ctacttgctt	atagacagat Tatctgtcta
3781	CGCTGAGATA GCGACTCTAT	GGTGCCTCAC	tgattaagca actaattcgt	ttggtaactg aaccattgac	TCAGACCAAG AGTCTGGTTC	tttactcata Aaatgagtat	tatactitag Atatgaaatc
3851	attgatttaa taactaaatt	aacticatit Tigaagtaaa	ttaatttaaa Aattaaattt	AGGATCTAGG TCCTAGATCC	TGAAGATCCT ACTTCTAGGA	ttttgataat Aaaactatta	CTCATGACCA GAGTACTGGT
3921	AAATCCCTTA TTTAGGGAAT		ACGTGAGTTT TCGTTCCACT TGCACTCAAA AGCAAGGTGA	Gagcgtcaga Ctcgcagtct	CCCCGTAGAA GGGGCATCTT	aagatcaaag ttctagtttc	gatcttcttg Ctagaagaac
3991	agatcctttt Tctaggaaa	tttctgcgcg aaagacgcgc	Taatctgctg Attagacgac	CTTGCAAACA GAACGTTTGT	aaaaaaccac ttttttggtg	CGCTACCAGC GCGATGGTCG	GGTGGTTTGT CCACCAAACA
4061.	TTGCCGGATC AACGGCCTAG	AAGAGCTACC TTCTCGATGG	AACTCTTTTT TTGAGAAAAA	CCGAAGGTAA GGCTTCCATT	CTGGCTTCAG GACCGAAGTC	CAGAGCGCAG GTCTCGCGTC	ataccaaata Tatggtttat
4131	CTGTCCTTCT GACAGGAAGA	AGTGTAGCCG TCACATCGGC	TAGTTAGGCC ACCACTTCAA ATCAATCCGG TGGTGAAGTT	accacttcaa tggtgaagtt	Gaacteteta Cttgagacat	GCACCGCCTA	Catacctcgc Gtatggagcg

FIG. 42G

4201	TCTGCTAATC AGACGATTAG	CTGTTACCAG GACAATGGTC	TGGCTGCTGC ACCGACGACG	CAGTGGCGAT GTCACCGCTA	AAGTCGTGTC TTCAGCACAG	TTACCGGGTT AATGGCCCAA	GGACTCAAGA CCTGAGTTCT
4271	CGATAGTTAC GCTATCAATG	CGGATAAGGC GCCTATTCCG		GGCTGAACGG CCGACTTGCC	GGGGTTCGTG CCCCAAGCAC	CACACAGCCC GTGTGTCGGG	agcttggagc Tcgaacctcg
4341	GAACGACCTA CTTGCTGGAT	Caccgaactg Gegetegac	agatacctac tctatggatg	AGCGTGAGCT TCGCACTCGA	atgagaaagc tactctttcg	GCCACGCTTC	CCGAAGGGAG GGCTTCCCTC
4411	AAAGGCGGAC TTTCCGCCTG	aggtatccgg Tccataggcc	Taagcggcag Attcgccgtc	GGTCGGAACA CCAGCCTTGT	GGAGAGCGCA CCTCTCGCGT	CGAGGGAGCT	TCCAGGGGGA AGGTCCCCCT
4481	AACGCCTGGT TTGCGGACCA	atctttatag Tagaaatatc	TCCTGTCGGG	TTTCGCCACC AAAGCGGTGG	TCTGACTTGA AGACTGAACT	GCGTCGALTT CGCAGCTAAA	ttgtgatgct Aacactacga
4551	COTCAGGGGG	GCGGAGCCTA	TGGAAAAACG ACCTTTTTGC	CCAGCAACGC GGTCGTTGCG	geccttttta ccggaaaat	CGGTTCCTGG GCCAAGGACC	CCTTTTGCTG GGRAAACGAC
4621	gccttttgct Cggaaaacga	Cacatotict Gegeacaaga	TTCCTGCGTT ATCCCCTGAT AAGGACGCAA TAGGGGACTA	atcccctgat Taggggacta	TCTGTGGATA AGACACCTAT	accgtattac Tggcataatg	CGCCTTTGAG GCGGAAACTC
4691	TGAGCTGATA ACTCGACTAT		CAGCCGAACG GTCGGCTTGC	ACCGAGCGCA TGGCTCGCGT	GCGAGTCAGT CGCTCAGTCA	GAGCGAGGAA CTCGCTCCTT	GCGGAAGAGC CGCCTTCTCG
4761	GCCCAATACG CGGGTTATGC		CAAACCGCCT CTCCCCGCGC GTTTGGCGGA GAGGGGCGCG	GTTGGCCGAT CAACCGGCTA	TCATTAATGC AGCTGGCACG AGTAATTACG TCGACCGTGC	AGCTGGCACG TCGACCGTGC	acagetitcc Tgtccaaagg
4831	CCACTGGAAA GCTGACCTTT		GCGGGCAGTG AGCGCAACGC AATTAATGTG CGCCCGTCAC TCGCGTTGCG TTAATTACAC	aattaatgig Ttaattacac	agttagctca Tcaatcgagt	CTCATTAGGC GAGTAATCCG	ACCCCAGGCT TGGGGTCCGA
4901	TTACACTUTA AATGTGAAAT		TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACGAAGGCCG AGCATACAAC ACACCTTAAC ACTCGCCTAT	TGTGGAATTG	TGAGCGGATA	ACAATTTCAC TGTTAAAGTG	ACAGGAAACA TGTCCTTTGT

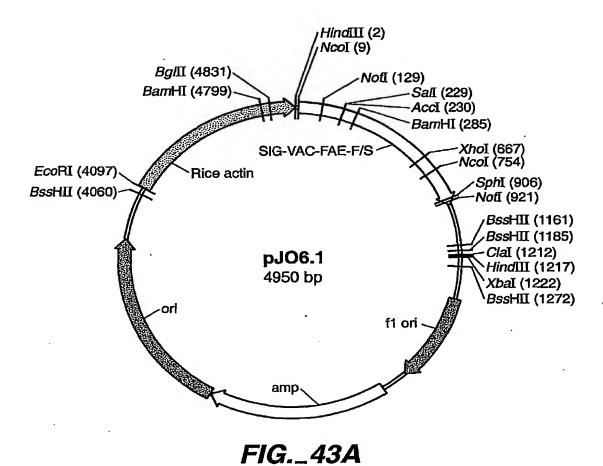
## 7/G.\_42H

NCOL

KpnI

GCTATGACCA TGATTACGCC AAGCGCGCAA TTAACCCTCA CTAAAGGGAA CAAAAGCTGG GTAC

BSSHII



SUBSTITUTE SHEET (RULE 26)

Hindill Mcol

りこうりょうじょう AGCGGCACGA CCGGTGCCGG CGGCAGCGGC GGCCACGGCC TCGCCGTGCT TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG CTCCTGGCGC CCGCGTCCTC AAGCTTACCA TGGCCCACGC <del>, i</del>

creececec eccesagere GACCGCGCGG CCGCCTCCAC CGGCCAGTGG GCCCGTCACC GAGGAGGAAG CGGCTGAGGT TGGGCTAGGC ACCCGATCCG GCCGACTCCA CICCICCITC AGCGGAGGAG TCGCCTCCTC

TGCCTACGCC TACCGGTGAT AGAGGGTTCG ACGGATGCGG TCTCCCAAGC TTTAGTCGAA ATGGCCACTA AGGCTTCTGG AGATGTCGGC AAATCAGCTT TCCGAAGACC TCTACAGCCG GCAGGGCATC

141

7

Sali

Acci

GACATTAACG CTGTAATTGC TTTAAATGTT AAGAGTTTGA TTCTCAAACT AAATTTACAA CTGGACACGT TGTAAGGCAG CTGATAATAG TTCCCTCTCT AAGGGAGAGA ACATTCCGTC GACTALTATC GACCTGTGCA

BamHI

GGCACTGGTA GTGATACGAA CACTATGCTT GCAGAAGGCA CCGTGACCAT CGTCTTCCGT CCGCGACGAC AGCAGCAAAG AAATAATCAC CTACCTAGGA GGCGCTGCTG TCGTCGTTTC TTTATTAGTG GATGGATCCT 281

TTGTGAAGTA AACACTTCAT TIACGITGCC AATGCAACGG ACCCTACCAC TGGGATGGTG GCCTTTCGAC CGGAAAGCTG TGTGGGAGTG ACACCCTCAC GATACTAACT CTATGATTGA TCTACAACTC AGATGTTGAG 351

CAGCAGGTTA GICGICCAAI GCTTGTCAAA CGAACAGTTT CAGGTCCTGG TTCAGCTCAG AAGTCGAGTC GTCCAGGACC ATGGGTCTCC TACCCAGAGG CACGGTGGAT ATTATATTGG TAATATAACC GTGCCACCTA 421

CACTCACTGC GTGAGTGACG TCCCTGGCGG AGGGACCGCC GGAGCCGCGG CCTCGGCGCC CCGGCCACKC GACTGGCACT GGCCGGTGMG CTGACCGTGA GGACTACGCG CCTGATGCGC GCCAGTATCC CGGTCATAGG 491

GCCGTTAGTC CGGCAATCAG AACCGCGCAG recrement decembers regardede trescent ACGACAACAT CCGCCTGTAC ACCTTCGGCG TCTGCGACAT GCGGGTCGAC AGACGCTGTA CGCCCAGCTG 561

-1G.\_43B

GACGCAGTAT TTCCGGGTCA CTGCGTCATA AAGGCCCAGT CGATGCCTTC CAAGCCTCGA GCCCAGATAC GCTACGGAAG GTTCGGAGCT CGGGTCTATG GCATGTACTT CGTACATGAA CGGAAGCGCA GCCTTCGCGT

GTGTAGAGTA GCCCATGGCG GCAGGGGTAC CCCCGGTGGA CCACGCCATC CCAAACCTGC GCTGCCGCGTTGGACG CTCATGCCAA GAGTACGGTT

CACATCTCAT CGGGTACCGC CGTCCCCATG GGGGCCACCT

GTGCTGTGAG CACGACACTC TACTTCACGT ATGAAGTGCA TGCACTGGGG ACGIGACCCC CACATTTGTC GTGTAAACAG GCGCCCAGAA CGCGGGTCTT

GATCCTTACA CTAGGAATGT

GACCTCGCAA

CTGGAGCGTT

771

701

631

Sphī

GACGAGCGGC

GCATGCACCT CGTACGTGGR

CITATTACGC GIGIGCIGAA TAAAACCCTA CIGCICGCCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT

Noti

CTGTCCCACA

CGGGTCCCGC

GCCCAGGGCG

841

CATTTGGCAA GTAAACCGTT AAAGAAGCAG ATCGTTCAAA TAGCAAGTTT TTTCTTCGTC GATGAGCTGT CTACTCGACA ACCACTGAAG TGGTGACTTC GGCCGCGGAA CCGGCGCCTT GGCCGGTCGC CCGGCCAGCG 911

TTGAATTACG AACTTAATGC TGATTATCAT ATAATTTCTG ACTAATAGTA TATTAAAGAC CCAGAACGCT GGTCTTGCGA TCCTGTTGCC AGGACAACGG ATTCTAACTT TAAGATTGAA ATTTCAAAGA TAAAGTTTCT 981

TTAGAGTCCC AATCTCAGGG CAAAAATACT GTTTTTATGA TGACGTTAIT TATGAGATGG ACTGCAATAA ATACTCTACC AATAATTAAC ATGTAATGCA TTATTAATTG TACATTACGT TTAAGCATGT AATTCGTACA 1051

Beghii BESHII

\*\*\*\*\*\*\*

GCAATTATAC ATTTAATACG CGATAGAAA CAAAATATAG CGCGCAAACT AGGATAAATT ATCGCGCGCG cgttaataig taaattaigc gctaicttit gittaataic gcgcgittga icctaittaa tagcgcgcg

CGCGCCGCCC

GCGAAAGAAG

このこれをいること

AATCCCAAGG

TTAGGGTTCC

Clar Hindir

GGATATCACT CCTATAGTGA GGCCACCTCG AGGTTAAGCG TCCAATTCGC CCGGTGGAGC CACAGTAGAT ACAATGATCT AGCTATTCGA AGATCTCGCC TGTTACTAGA TCGATAAGCT TCTAGAGCGG GTGTCATCTA 1191

CGTTACCCAA CGTGACTGGG AAAACCCTGG TEGCCGTCGT TTTACAACGT GCGCGCTCAC Besell GTCGTATTAC

GCAATGGGTT CGCGCGAGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC TTTTGGGGACC CAGCATAATG 1261

TGGCTAGCGG ACCGATCGCC AGAGGCCCGC TCTCCGGGCG CATTATCGCT GTAATAGCGA GCCAGCTGGC CGGTCGACCG TCCCCCTTTC AACGTCGTGT AGGGGGAAAG TTGCAGCACA GAATTAGCGG CTTAATCGCC

ອລອວອອວອວອ GTTGCGCAGC

GGCGCATTAA CCGCGTAATT CGGGACATCG GCCCTGTAGC AATGGGACGC TTACCCTGCG CTGAATGGCG GACTTACCGC CAACGCGTCG

CCCCTCTT GGCGAGGAAA GCCCTAGCGC CGGGATCGCG ACTTGCCAGC TGAACGGTCG TGACCGCTAC ACTGGCGATG ACGCGCAGCG TGCGCGTCGC

GGGGCTCCCT CCCCGAGGGA CTCTAAATCG GAGATTTAGC CCCCGTCAAG GGGCCAGTTC CGCCGGCTTT GCGCCCGAAA TCGCCACGTT AGCGGTGCAA GGAAGGAAAG CCTTCCTTTC

CACCCGGTAG Gradaccarc CCAAGTGCAT GGTTCACGTA TTAGGGTGAT AATCCCACTA TTTTGAACT AAAAACTTGA CTCGACCCCA GAGCTGGGGT TTTACGGCAC AAATGCCGTG GATTTAGTGC CTAAATCACG 1611

CTTGTTCCAA GAACAAGGTT TATCACCTGA ATAGTGGACT ACGUTCTUTA TGCAAGAAAT GTTGGAGTCC CAACCTCAGG GCCCTTTGAC CGGGAAACTG ACGGTTTTTC TGCCAAAAAG GCCCTGATAG CGGGACTATC 1681

ATTTCGGCCT TAAAGCCGGA GATTTTGCCG CTAAAACGGC TAAATATTCC ATTTATAGG TATTCTTTG ATAGAGCCAG ATAAGAAAC TATCTCGGTC GTGAGTTGGG CACTCAACCC ACTGGAACAA TGACCTTGTT 1751

GCGAATGTTA TTTATATT AAAATATTAA GAATTTTAAC CTTAAAATTG TTTACTCGAC TAAATTGTTT TTAAATTGCG ATTTAACAAA AATTTTAACGC AAATGAGCTG TAACCAATTT ATTGGTTAAA 1821

۶.

SUBSTITUTE SHEET (RULE 26)

CTTCCCAACA

1401

1331

GAAGGGTTGT

TGTGGTGGTT ACACCACCAA

1471

1891	TTAGGTGGCA AATCCACCGT	CTTTTCGGGG	AAATGTGCGC TTTACACGCG	GGAACCCCTA CCTTGGGGAT	tttgittait tttctaaata Aaacaaataa aaagatttat	tttctaarta Aaagatttat	Cattcaaata Gtaagtttat
1961	TGTATCCGCT ACATAGGCGA	Catcagacaa Gtactctgtt	Taaccctgat Aftgggacta	aaatgcttca tttacgaagt	ataatatega Tattataact	aaaaggaaga ttttccttct	gtatgagtat Catactcata
2031	TCAACATTTC AGTTGTAAAG	CGTGTCGCCC	TTATTCCCTT AATAAGGGAA	TTTTGCGGCA AAAACGCCGT	TTTTGCCTTC AAAACGGAAG	CTGTTTTTGC GACAAAAACG	TCACCCAGAA AGTGGGTCTT
2101	accetegtea Tecgaccact	aagtaaaaga Ttcattttct	tgctgaagat acgacttcta	CAGTTGGGTG GTCAACCCAC	CACGAGTGGG GTGCTCACCC	ttacatcgaa aatgtagctt	CTGGATCTCA GACCTAGAGT
2171	ACAGCGGTAA TGTCGCCATT	GATCCTTGAG CTAGGAACTC	AGTTTTCGCC TCAAAAGCGG	CCGAAGAACG GGCTTCTTGC	ttttccaatg Aaaaggttac	atgagcactt Tactcgtgaa	ttaaagttct aatttcaaga
2241	GCTATGTGGC CGATACACCG	GCGGTATTAT CGCCATAATA	CCCGTATTGA GGGCATAACT	CGCCGGGCAA	GAGCAACTCG CTCGTTGAGC	GTCGCCGCAT	acactatitct tgtgataaga
2311	CAGAATGACT GTCTTACTGA	TGGTTGAGTA ACCAACTCAT	CTCACCAGTC	acagaaaagc Tgtcttttcg	atcttacgga tagaatgcct	TGGCATGACA ACCGTACTGT	GTAAGAGAAT CATTCTCTTA
2381	Tatgcagtgc Atacgtcacg	TGCCATAACC ACGGTATTGG	ATGAGTGATA TACTCACTAT	ACACTGCGGC TGTGACGCCG	CAACTTACTT GTTGAATGAA	Ctgacaacga Gactgttgct	TCGGAGGACC
2451	GAAGGAGCTA CTTCCTCGAT	ACCGCTTTTT TGGCGAAAAA	TGCACAACAT ACGTGTTGTA	GGGGGATCAT CCCCCTAGTA	GTAACTCGCC CATTGAGCGG	TTGATCGTTG AACTAGCAAC	GGAACCGGAG CCTTGGCCTC
2521	Ctgaatgaag gacttacttc	CCATACCAAA GGTATGGTTT	CGACGAGCGT GCTGCTCGCA	GACACCACGA	TGCCTGTAGC	aatggcaaca Ttaccgttgt	acgttgcgca Tgcaacgcgt
2591	aactattaac Ttgataattg	TGGCGAACTA	CTTACTCTAG GAATGAGATC	CTTCCCGGCA	acaattaata Tgttaattat	Gactggatgg Ctgacctacc	aggcggataa TCCGCCTATT
2661	agttgcagga Tcaacgtcct	CCACTTCTGC GGTGAAGACG	GCTCGGCCCT	TCCGGCTGGC	TGGTTTATTG ACCAAATAAC	CTGATAAATC GACTATTTAG	TGGAGCCGGT ACCTCGGCCA

CACCOTGGGT C		CTCGCGGTAT GAGCGCCATA	CATTGCAGCA GTAACGTCGT	CTGGGGCCAG GACCCCGGTC	atggtaagcc taccattcgg	CTCCCGTATC GAGGGCATAG	gtagttatct Catcaataga
ACACGACGGG GAGTCAGGCA 1	GAGTCAGGCA	Left Tr.	actatggatg Tgatacctac	aacgaaatag Ttgctttatc	acagatogot Tgtctagoga	gagataggtg Ctctatccac	CCTCACTGAT GGAGTGACTA
TAAGCATTGG TAACTGTCAG A( ATTCGTAACC ATTGACAGTC T(	Taactgtcag Attgacagtc	ĂH	accaagtita Tggttcaaat	CTCATATATA GAGTATATAT	Ctttagattg Gaaatctaac	atttaaact Taaatttiga	tcatititaa Agtaaaaatt
TTTAAAAGGA TCTAGGTGAA G2 AAATTTTCCT AGATCCACTT C?	TCTAGGTGAA AGATCCACTT	පී වි	gatccttttt Ctaggaaaa	gataatctca Ctattagagt	TGACCAAAAT ACTGGTTTTA	CCCTTAACGT GGGAATTGCA	GAGTTTTCGT CTCAAAAGCA
TCCACTGAGC GTCAGACCCC G AGGTGACTCG CAGTCTGGGG C	GTCAGACCCC CAGTCTGGGG	<b>5</b> 5	GTAGAAAAGA CATCTTTTCT	TCAAAGGATC AGTTTCCTAG	ttcttgagat Aagaactcta	CCTTTTTTTC GGAAAAAAAG	tgcgcgtaat acgcgcatta
CTGCTGCTTG CAAACAAAA AAGACGAAGAAC GTTTGTTTTT T	Caaacaaaaa Gittgittit		aaccaccgct ttggtggcga	ACCAGCGGTG TGGTCGCCAC	gittgittgc Caaacaaacg	CGGATCAAGA GCCTAGTTCT	GCTACCAACT CGATGGTTGA
CTTTTTCCGA AGGTAACTGG COGAAAAAGGCT TCCATTGACC G	AGGTAACTGG TCCATTGACC		CTTCAGCAGA GAAGTCGTCT	GCGCAGATAC CGCGTCTATG	Caaatactgt Gettatgaca	CCTTCTAGTG GGAAGATCAC	tagccgtagt atcggcatca
TAGGCCACCA CTTCAAGAAC T	CTTCAAGAAC GAAGTTCTTG		TCTGTAGCAC AGACATCGTG	CGCCTACATA GCGGATGTAT	CCTCGCTCTG	Ctaatcctgt Gattaggaca	TACCAGTGGC ATGGTCACCG
TGCTGCCAGT GGCGATAAGT CGTGTCTTAC ACGACGGTCA CGGCTATTCA GCACAGAATG		บิษั	CGTCTCTTAC GCACAGAATG		CGGGTTGGAC TCAAGACGAT GCCCAACCTG AGTTCTGCTA	agttaccega Tcaatggcct	taaggcgcag attccgcgtc
CGGTCGGGCT GAACGGGGGG T	GAACGGGGGG		TTCGTGCACA AAGCACGTGT	CAGCCCAGCT	TGGAGCGAAC ACCTCGCTTG	GACCTACACC CTGGATGTGG	gaactgagat Cttgactcta
ACCTACAGCG TGAGCTATGA G TGGATGTCGC ACTCGATACT C	tgagctatga actcgatact		graagcgcca Ctttcgcggt	CGCTTCCCGA GCGAAGGGCT	CCCTTCCCGA AGGGAGAAAG GCGAAGGGCT TCCCTCTTTC	GCGGACAGGT CGCCTGTCCA	atccggtarg trggccattc
CGGCAGGGTC GGAACAGGAG A GCCGTCCTC CTTGTCCTC T	GGAACAGGAG CCTTGTCCTC		agcgcacgag Tcgcgtgctc	GGAGCTTCCA CCTCGAAGGT	GGGGGAAACG CCTGGTATCT CCCCCTTTGC GGACCATAGA	CCTGGTATCT GGACCATAGA	ttatagecct aatatcagga

FIG.\_431

3571	GTCGGGTTTC	GTCGGGTTTC GCCACCTCTG CAGCCCAAAG CGGTGGAGAC	ACTTGAGCGT CGATTTTTGT TGAACTCGCA GCTAAAACA	Cgattttttgt Gctaaaaga	gatectcetc Ctacgagcag	AGGGGGGCGG TCCCCCCGCC	agcctatgga tcggatacct
3641	AAAACGCCAG TTTTGCGGTC		tttttacggt aaaaatgcca	CAACGCGGCC TTTTTACGGT TCCTGGCCTT TTGCTGGCCT TTTGCTCACA GTTGCGCCGG AAAAATGCCA AGGACCGGAA AACGACCGGA AAACGAGTGT	TTGCTGGCCT AACGACCGGA	TTGCTGGCCT TTTGCTCACA AACGACCGGA AAACGAGTGT	tgitciticc Acaagaaagg
3711	TGCGTTATCC ACGCAATAGG	TGCGTTATCC CCTGATTCTG ACGCAATAGG GGACTAAGAC	TGGATAACCG ACCTATTGGC	Tattaccecc Ataateccee		TITICAGTGAG CTGATACCGC TCGCCGCAGC AAACTCACTC GACTATGGCG AGCGGCGTCG	TCGCCGCAGC
3781	CGAACGACCG	AGCGCAGCGA TCGCGTCGCT	GTCAGTGAGC CAGTCACTCG	GAGGAAGCGG	AAGAGCGCCC TTCTCGCGGG	aatacgcaaa Ttatgcgttt	CCGCCTCTCC
3851	CCGCGCGTTG	GCCGATTCAT CGGCTAAGTA	GCCGATTCAT TAATGCAGCT GGCACGACAG CGGCTAAGTA ATTACGTCGA CCGTGCTGTC	GGCACGACAG CCGTGCTGTC	<i>GTTTCCCGAC</i> CAAAGGGCTG	TGGAAAGCGG ACCTTTCGCC	GCAGTGAGCG CGTCACTCGC
3921	Caacgcaatt Gttgcgttaa		AATGTGAGTT AGCTCACTCA TTACACTCAA TCGAGTGAGT	ttaggcaccc aatccgtggg	Caggctttac Gtccgaaatg	actitiatget Tgaaatacga	TCCGGCTCGT AGGCCGAGCA
3991	Atgitgigi Tacaacacac	ATGTTGTGTG GAATTGTGAG TACAACACAC CTTAACACTC	CGGATAACAA GCCTATTGTT		gaaacagcta Ctttgtcgat	TGACCATGAT ACTGGTACTA	BSBELI ~~ TACGCCAAGC ATGCGGTTCG
	Basell			ECORI			
4061	GCGCAATTAA CGCGTTAATT		agggaacaaa TCCCTTGTTT	CCCTCACTAA AGGGAACAAA AGCTGGAATT CCACAATGAA GGGAGTGAIT TCCCTTGITT TCGACCTIAA GGTGTTACTI	CCACAATGAA GGTGTTACTT	CAATAATAAG ATTAAAATAG GTTATTATTC TAATTTTATC	attaaaatag taattttatc
4131	CTTGCCCCCG	CTTGCCCCCG TTGCAGCGAT GAACGGGGGC AACGTCGCTA	gggtatttt CCCataaaa	tctagtaaaa Agatcatttt	traragatra Attitctatt	acttagactc Tgaatctgag	aaacattta Ttttgtaaat
4201	Caaaaacaac Gittititgitg	CAAAAACAAC CCCTAAAGTC CTAAAGCCCA AAGTGCTATG GTTTTTGTTG GGGATTTCAG GATTTCGGGT TTCACGATAC	CTAAAGCCCA GATTTCGGGT	aagtgctatg ttcacgatac	Cacgatccat Gtgctaggta	CACGATCCAT AGCAAGCCCA GCCCAACCCA GTGCTAGGTA TCGTTCGGGT CGGGTTGGGT	GCCCAACCCA CGGGTTGGGT

# FIG.\_43G

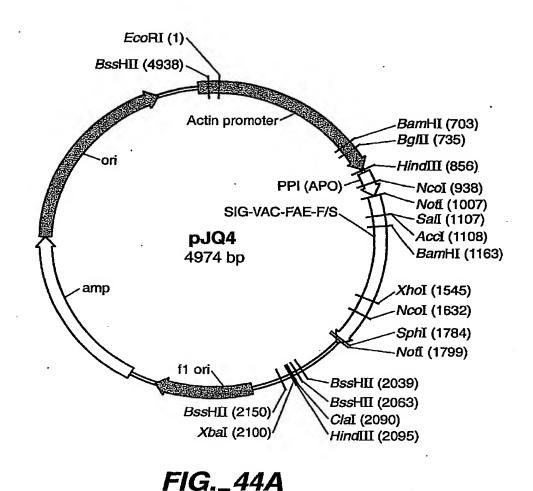
H.A.	ည္က စ္ဌ	ប្តូច្ច	ပ္ကစ္	e a	ខ្លួម	0 U	H * 44	T.
CACCGTGAGT GTGGCACTCA	gaaaaacagc Citititigicg	ರಾದ್ಯದ್ದು	TCCCCCCAAC AGGGGGGTTG	cccrccccr ggaagggaa	CTCGGTCTCG	CCTCCCGCGG	Bglli CTCGGATGTA GAGCCTACAT	agtyptycty Tcaaaaagaa
5 F		9 8 5 8				တ္တည တပ		
CCGGCACTAT CACCGTGAGT GGCCGTGATA GTGGCACTCA	aaaagaaaaa Ttttctttt	CGACGAGGCC	TCTCCTCCCA Agaggagggt	AGCTCCTCCC	tttttttcgt aaaaaaagca	rccarcccc agccarcccc	gaatggggtt Cttaccccga	gttcatcggt caagtagcca
GTCTCCACCC	agaaagaaa Tctttctttt	TCGCGAGCAG CGACGAGGCC CGGCCCTCCC	TACCCCCCC TCTCCTCCCA ATGGGGGGGGGGGGGGGGG	CCGGACGACG AGCTCCTCCC GGCCTGCTC TCGAGGAGGG	<b>TTCTCCGTTT</b> AAGAGGCAAA	GTCGCCCAGA TCGGTGCGCG CAGCGGGTCT AGCCACGCGC	HI TCCTCGCGGG AGGAGCGCCC	CCTCAGCATT GTTCATCGGT GGAGTCGTAA CAAGTAGCCA
CTGGCAAATA GACCGTTTAT	aaaaaaaa Tiitiiiiii	GCGAGGAGGA TCGCGAGCAG CGCTCCTCT AGCGCTCGTC	ACTATATACA TGATATATGT	CCCCTCGCTG	CGGTAACCAC CCCGCCCCTC TCCTCTTTCT TTCTCCGTTT TTTTTTCGT GCCATTGGTG GGGCGGGAG AGGAGAAGA AAGAGGCAAA AAAAAAAGCA	GAGCGGCTTC	Bambi crescercre ceescercas reseccesa recreseses saccerase seccesece	aatttgaatc Ttaaacttag
GTGCAGCCAA CACGTCGGTT	TCGCAGCCAA AGCGTCGGTT		CCCCATCGCC	CACCTCCTCC GTGGAGGAGG	CCCCCCTC	GGGTGGGCGA CCCACCCGCT	CGGGCGTGAG	CTTTCTTCTT TTTGTGGTAG
ACCCAACCCA ACCCACCCCA GTGCAGCCAA CTGGCAAATA TGGGTTGGGT TGGGTGGGGT CACGTCGGTT GACCGTTTAT	ACCGCACGTC TGGCGTGCAG	AGGTGGGTCC GGGTCGTGGG GGCCGGAAAA TCCACCCAGG CCCAGCACCC CCGGCCTTTT	aagaaacgcc ttctttgcgg	CCACCACCAC	CGGTAACCAC GCCATTGGTG	TTGGTAGTTT AACCATCAAA	CTGGCGTCTC	CTTTCTTCTT GAAAGAAGAA
ACCCAACCCA TGGGTTGGGT	TOTCCGCACC	AGGTGGGTCC	TCCGCTTCCA	CCTACCACCA	0099099099	ATCTTTGGCC TAGAAACCGG	GATCTCGCGG CTAGAGCGCC	Bglii carctictit ctagaagaa
4271	4341	4411	4481	4551	4621	4691	4761	4831

FIG.\_43H

TTCATGATTT GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGGTAG AAGTACTAAA CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCCATC

4901

SUBSTITUTE SHEET (RULE 26)



SUBSTITUTE SHEET (RULE 26)

ECORI

러	aattccacaa Ttaaggtgtt	TGAACAATAA ACTTGTTATT	TAAGATTAAA ATAGCTTGCC CCCGTTGCAG ATTCTAATTT TATCGAACGG GGGCAACGTC	atagcttgcc tatcgaacgg		CGATGGGTAT TTTTTCTAGT GCTACCCATA AAAAAGATCA	tttttctagt Aaaaagatca
71	aaaatraaag Ttttaitttc	ataaacttag tatttgaatc	actcaaaaca tgagitittgi	titacaaaaa aaatgitiit	CAACCCCTAA GTTGGGGATT	AGTCCTAAAG TCAGGATTTC	CCCAAAGTGC GGGTTTCACG
141	TATGCACGAT ATACGTGCTA	CCATAGCAAG GGTATCGTTC	CCCAGCCCAA	CCCAACCCAA	CCCAACCCAC	CCCAGTGCAG GGGTCACGTC	CCAACTGGCA GGTTGACCGT
211	aatagtctcc Ttatcagagg	ACCCCCGGCA TGGGGGCCGT	CTATCACCGT	GAGTTGTCCG	CACCACCGCA	CGTCTCGCAG GCAGAGCGTC	CCAAAAAAA GGTTTTTTTT
281	aaaaagaaag Tititcitic	aaaaaaaga Ttttttttct	aaaagaaaa Tittcctttt	CAGCAGGTGG GTCGTCCACC	GTCCGGGTCG	TGGGGGCCGG ACCCCCGGCC	aaaagcgagg Tittcgctcc
351	AGGATCGCGA TCCTAGCGCT	GCAGCGACGA	<u> </u>	TCCCTCCGCT	TCCAAAGAAA AGGTTTCTTT	CGCCCCCCAT	CGCCACTATA GCGGTGATAT
421	TACATACCCC ATGTATGGGG	CCCCTCTCCT	CCCATCCCCC GGGTAGGGGG	CAACCCTACC GTTGGGATGG	ACCACCACCA TGGTGGTGGT	CCACCACCTC GGTGGTGGAG	CTCCCCCTC
491	GCTGCCGGAC	GACGAGCTCC CTGCTCGAGG	TCCCCCTTC CCCTCCGCCG	CCCTCCGCCG	CCGCCGGTAA	CCACCCGCC	CCTCTCCTCT GGAGAGGAGA
561	TTCTTTCTCC AAGAAAGAGG	GTTTTTTTT Caaaaaaaa	TCGTCTCGGT AGCAGAGCCA	CTCGATCTTT GAGCTAGAAA	GGCCTTGGTA CCGGAACCAT	GTTTGGGTGG CAAACCCACC	GCGAGAGCGG
631	CTTCGTCGCC GAAGCAGCGG	CAGATCGGTG GTCTAGCCAC	CGCGGGAGGG GCGGGATCTC GCGCCCTCCC CGCCCTAGAG	GCGCGATCTC	GCGGCTGGCG	TCTCCGGGCG AGAGGCCCGC	TGAGTCGGCC
701	Bamhl CGGATCCTCG GCCTAGGAGC	CGGGGAATGG GCCCCTTACC	GGCTCTCGGA CCGAGAGCCT	Bglii """" TGTAGATCTT ACATCTAGAA	CTTTCTTTCT GAAAGAAAGA	tctttttgt agararacac	gtagaatttg Catcttaaac

FIG.\_44B

CGTGCGGAGC	GCACGCCTCG
AATGCAGCCT	Itagggagto gtarcaagta gccatcaaaa agaaaagtac taaacactgt ttacgtcgga gcacgcctcg
ATTTGTGACA	TAAACACTGT
TCTTTTCATG	AGAAAAGTAC
CGGTAGITIT	GCCATCAAAA
CATTGTTCAT	GTAACAAGTA
771 AATCCCTCAG CATTGTTCAT CGGTAGTTTT TCTTTTCATG ATTGTGACA AATGCAGCCT CGTGCGGAGC	TTAGGGAGTC
771	

### Hindri

TCCTGATCGT CCASTTGAAG CASCGGATGG AGGACTAGCA GGTSAACTTC GTSGCCTACC GRAGAAGCTT ACMATGGMCG TGCACAAGGA TGKTACCKGC ACGTGTTCCT CATCTTCGAA TTTTTGTAG AAAAAACATC 841

CGAGTGCGGC GCTCACGCCG SCTCGGCCTC CTCTTGCTCG TSTCCGCCAT GGAGCACGTG GACGCCAAGG CCTGCACCCK SGAGCCGGAG GAGAACGAGC ASAGGCGGTA CCTCGTGCAC CTGCGGTTCC GGACGTGGGW GACGCCAAGG TSTCCGCCAT GGAGCACGTG

### Noti

22222222

CCCGGCGGCC GCCTCCACGC AGGCCATCTC CGAAGACCTC TACAGCCGTT GGGCCGCGG CGCAGGTGCG TCCCGTAGAG GCTTCTGGAG ATGTCGGCAA TTGGAGCCGA AGCCGTAGAC TCGGCATCTG AACCTCGGCT

### \*\*\*\*\*\* SalI

Acci

APPCCGTCGA CTATTATCAA GATAATAGTT \*\*\*\*\*\*\* TAAGGCAGCT GGACACGTTG CCTGTGCAAC TAGTCGAAAT GGCCACTATC TCCCAAGCTG CCTACGCCGA ATCAGCTTTA CCGGTGATAG AGGGTTCGAC GGATGCGGCT

### BamHI

CAGCAAAGAA GCGACGACAG TGGATCCTCC CATTAACGGA CTCAAACTGA ATTACAATT

CGCTGCTGTC GTCGTTTCTT ACCUTCACGC TACTAACTAC GTAATTGCCT ACCTAGGAGG TACAACTCGA GATACGAATC GAGTTTGACT CACTGGTAGT TCTTCCGTGG AGAAGGCACC TAAATGTTAA GGGAGAGAAA CCCTCTCTTT ATAATCACCG 1121 1191

GGGTCTCCGT CCCAGAGGCA GCCACCTATA ATATAACCTA TATATTGGAT CGGTGGATAT ACGITGCCAA CACTICAIGI TGCAACGGTT GTGAAGTACA GGATGGTGTT CCTACCACAA CTTTCGACAC GAAAGCTGTG 1261

Greaccarca crargetrag argregaecr argarreare regeagreeg

TATTAGTGGC

981

1051

1331	CCAGGACCAA GGTCCTGGTT	GTCGAGTCGC	ttgtcaaaca aacagtttgt	GCAGGTTAGC CGTCCAATCG	CAGTATCCGG GTCATAGGCC	CAGTATCCGG ACTACGCGCT GTCATAGGCC TGATGCGCGA	GACCGTGACC CTGGCACTGG
1401	GGGCTGMGGG	TCGGCGCCTC	CCTGGCGGCA	CTCACTGCCG	CCCAGCTGTC TGCGACATAC GGGTCGACAG ACGCTGTATG	TGCGACATAC ACGCTGTATG	Gacaacatcc Ctgttgtagg
1471	GCCTGTACAC		CTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTCGCGTCG TACATGAACG GAAGCCGCTT GGCGCGTCGC CGTTAGTCCG GAAGCGCAGC ATGTACTTGC	GCAATCAGGC CGTTAGTCCG	CTTCGCGTCG TACATGAACG GAAGCGCAGC ATGTACTTGC	TACATGAACG ATGTACTTGC	ATGCCTTCCA TACGGAAGGT
1541	Xhoi ZZZZZZZ AGCCTCGAGC TCGGAGCTCG	CCAGATACGA GGTCTATGCT	CGCAGTATTT GCGTCATAAA	CCGGGTCACT	CATGCCAACG ACGGCATCCC GTACGGTTGC TGCCGTAGGG	ACGGCATCCC TGCCGTAGGG	AAACCTGCCC T'TTGGACGGG
1611	CCGGTGGAGC GGCCACCTCG	AGGGGTACGC TCCCCATGCG	Ncol CCATGGCGGT GGTACCGCCA	GTAGAGTACT CATCTCATGA	GGAGCGTTGA CCTCGCAACT	TCCTTACAGC	GCCCAGAACA CGGGTCTTGT
1681	CATTTGTCTG GTAAACAGAC	CACTGGGGAT GTGACCCCTA	GAAGTGCAGT CTTCACGTCA	GCTGTGAGGC	CCAGGGGGA CAGGGTGTGA GGTCCCGCCT GTCCCACACT	CCAGGGCGGA CAGGGTGTGA GGTCCCGCCT GTCCCACACT	ataatgcgca tattacgcgt
1751	CACGACTTAT GTGCTGAATA	titigggatga Aaaccctact	By COAGCGGCGC GCTCGCGCG	Sphi ccarctee ce taceteeacc	Noti CCGGTCGCGG CCGCGGAAAC GGCCAGCGCC GGCGCCTTFG	CCCCGGAAAC CCCCCCTTTG	CACTGAAGGA GTGACTTCCT
1821	TGAGCTGTAA ACTCGACATT	AGAAGCAGAT TCTTCGTCTA	CGTTCAAACA GCAAGTTTGT		aagiticita Ticaaagaat	AGATTGAATC TCTAACTTAG	CTGTTGCCGG
1891	rcttgcgatg agaacgctac		ATTATCATAT AATTTCTGTT GAATTACGTT AAGCATGTAA TAATTAACAT GTAATGCATG TAATAGTATA TTAAAGACAA CTTAATGCAA TTCGTACATT ATTAATTGTA CATTACGTAC	gaattacgit Citaaigcaa	aagcatgtaa Ttcgtacatt	taattaacat Attaattgea	GTAATGCATG CATTACGTAC
1961	ACGTTATTTA TGCAATAAAT	TGAGATGGGT ACTCTACCCA	TGAGATGGGT TTTTATGATT AGAGTCCCGC AATTATACAT TTAATACGCG ACTCTACCCA AAATTATGCGC TTAATATGTA AATTATGCGC	AGAGTCCCGC TCTCAGGGCG	AATTATACAT TTAATACGCG TTAATATGTA AATTATGCGC	TTAATACGCG AATTATGCGC	atagaaaaca Tatcttttgt

SUBSTITUTE SHEET (RULE 26)

нтн оө	En ct	ប្រុ	đ H	ប្រ	೮ಠ	e e	E+ c¢	<b>4</b> E	d H
Xbai Hindiii PAGCTTC	GTT	FTCG	3CGA	FACA	FTTC	SGLT	SACG	GAGA	AAAA I'I'I'I
: 55	GCCGTCGTTT	CCCCTTTCGC	gaatggcgaa Cttaccgctt	accectacac Tggcgatgtg	CCGGCTTTCC	CGACCCCAAA GCTGGGGTTT	CCTTTGACGT GGAAACTGCA	TCTCGGTCTA AGAGCCAGAT	ttaacaaaaa aattgittit
Clai frogu									PA TA TA
FAGA:	<b>10.</b>	PACA	PTCG(	1565 1565	CGTT	3CAC Cord	TTTC	ACCC TGGG	SCTG
Cle ZZZ TTACTAGATC	BSBHII  CGTATTACGC GCGCTCACTG GCATAATGCG CGCGAGTGAC	GCAGCACATC	CGATCGCCCT TCCCAACAGT TGCGCAGCCT GCTAGCGGGA AGGGTTGTCA ACGCGTCGGA	GCGCAGCGTG	GCCACGTTCG	TTTAGTGCTT TACGGCACCT AAATCACGAA ATGCCGTGGA	GGTTTTTCGC CCAAAAAGCG	CTCAACCCTA GAGTTGGGAT	atgagctgat Tactcgacta
ATG	Babhii CGC GCG GCG CGG		AGT			GAA			
GTCATCTATG CAGTAGATAC	BBB  CGTATTACGC GCATAATGCG	TTACCCAACT TAATCGCCTT AATGGGTTGA ATTAGCGGAA	TCCCAACAGT	GCGGCGGGTG TGGTGGTTAC CGCCGCCCAC ACCACCAATG	TTCCTTTCTC AAGGAAAGAG	tttagtgett Aaatcacgaa	CCTGATAGAC GGACTATCTG	TGGAACAACA ACCTTGTTGT	TTCGGCCTAT TGGTTAAAA AAGCCGGATA ACCAATTTT
GATO CASO		ALT	77.	TGG ACC	TTC		25 e	TGG ACC	TGG.
BSSHII CGCGCGGGT CGCGCGGGT	tatagtgagt atatcactca	TTACCCAACT AATGGGTTGA	CGATCGCCCT GCTAGCGGGA	GGTG	CTTTCTTCCC Gaaagaaggg	aggottccaa TCCCAAGGCT	GGGCCATCGC CCCGGTAGCG	TGTTCCAAAC ACAAGGTTTG	CTAT
Bashii Gegega Tinasa Gegega	TAGT	ACCC	ATCG	ညည	TTCT	CCAA	GCCA	TTCC	ရှိသည်။ သူ့ရရှိသည်။
		o o			5 G		99 55 E+ 44	T TG	T A
ATTA TAAT	ည္သမ္မာ	TGGC	CGCA	TAAG	TTTC	CCTT	GTAG	ACTC	CCGA GGCT
GATAAATTAT CTATTAATA	Caattcgccc Getaagcggg	AACCCTGGCG TTGGGACCGC	AGGCCCGCAC TCCGGGCGTG	CGCATTAAGC GCGTAATTCG	GCTCCTTTCG CTTTCTTCCC TTCCTTTCTC CGAGGAAAGC GAAAGAAGAG AAGGAAAAAA	GGCTCCCTTT CCGAGGGAAA	TTCACGTAGT GGGCCATCGC AAGTGCATCA CCCGGTAGCG	AGTGGACTCT TGTTCCAAAC TGGAACAACA TCACCTGAGA ACAAGGTTTG ACCTTGTTGT	t't't'gccgat aaaacgcta
aact Ttga	GAGC	TGGG	GCGA	TAGC	ညည	ATCG	TGAT	GTTCTTTAAT CAAGAAATTA	ttataaggga Aatattccct
BESHII ~~~~~~ AAATATAGCG CGCAAACTAG TTTATATCGC GCGTTTGATC	GGTGGAGCTC	TGACTGGGAA ACTGACCCTT	aatagcgaag ttatcgcttc	CCTGTAGCGG	CCTAGCGCCC GGATCGCGGG	CTAAATCGGG GATTTAGCCC	AAACTTGATT AGGGTGATGG TTTGAACTAA TCCCACTACC	GTTCTTTAAT CAAGAAATTA	ttataaggga aatattccct
Bashii GCG CGC					ည္သမ္သ	GCT	TAA	TGGAGTCCAC ACCTCAGGTG	GAT
TATA	xdai ~~~~ Tagagcggcc Atctcgccgg	tacaacgtcg atgitgcagc	CAGCTGGCGT GTCGACCGCA	TGGGACGCGC	TTGCCAGCGC AACGGTCGCG	CCGTCAAGCT GGCAGTTCGA	aaacttgatt Tttgaactaa	TGGAGTCCAC ACCTCAGGTG	tectetegae Aagaaaacea
AAA TIT	xbai reca	TAC	CAG	TGG	TTG	င်င် ဝင်င်	AAA	TGG	TTC
2031	2101	2171	2241	2311	2381	2451	2521	2591	2661
N	0	0	N	CA.	0	N	N	0	0

2731	TTTAACGCGA	attttaacaa	aatattaacg	CTTACAATTT	AGGTGGCACT	TTTCGGGGAA	atgrececes
	AAATTGCGCT	taaaattgit	Ttataattgc	GAATGTTAAA	TCCACCGTGA	AAAGCCCCTT	Tacacecec
2801	AACCCCTATT	tgtttattt	TCTAAATACA	ttcaaatatg	TATCCGCTCA	TGAGACAATA	accetgataa
	TTGGGGATAA	Acaaataaaa	AGATTTATGT	aagtttatac	ATAGGCGAGT	ACTCTGTTAT	Tgggactatt
2871	atgcttcaat	aatattgaaa	aaggaagagt	atgagtattc	aacatttccg	TGTCGCCCTT	attccctttt
	tacgaagtta	Ttataacttt	ttccttctca	Tactcataag	ttgtaaaggc	ACAGCGGGAA	Taagggaaaa
2941	TTGCGGCATT	TTGCCTTCCT	gtttttgctc	acccagaaac	GCTGGTGAAA	gtaaaagatg	Ctgaagatca
	AACGCCGTAA	AACGGAAGGA	Caaaaacgag	Tgggtctttg	CGACCACTTT	Cattttctac	Gacttctagt
3011	GTTGGGTGCA	CGAGTGGGTT GCTCACCCAA	ACATCGAACT TGTAGCTTGA	GGATCTCAAC CCTAGAGTTG	agcggtaaga Tcgccattct	TCCTTGAGAG AGGAACTCTC	TTTTCGCCCC AAAAGCGGGG
3081	gaagaacgtt	TTCCAATGAT	gagcactttt	aaagttctgc	Tatgtggggg	GGTATTATCC	CGTATTGACG
	Cttcttgcaa	AAGGTTACTA	Ctcgtgaaaa	tttcaagacg	Atacaccgcg	CCATAATAGG	GCATAACTGC
3151	CCGGGCAAGA	GCAACTCGGT CGTTGAGCCA	CGCCGCATAC GCGGCGTATG	actattctca Tgataagagt	GAATGACTTG CTTACTGAAC	GTTGAGTACT CAACTCATGA	CACCAGTCAC GTGGTCAGTG
3221	AGAAAAGCAT TCTTTTCGTA	CTTACGGATG GAATGCCTAC	GCATGACAGT	aagagaatta TTCTCTTAAT	TGCAGTGCTG ACGTCACGAC	CCATAACCAT GGTATTGGTA	Gagtgataac Ctcactattg
3291	ACTGCGGCCA	ACTTACTTCT	GACAACGATC	GGAGGACCGA	AGGAGCTAAC	CGCTTTTTTG	CACAACATGG
	TGACGCCGGT	TGAATGAAGA	CTGTTGCTAG	CCTCCTGGCT	TCCTCGATTG	GCGAAAAAAC	GTGTTGTACC
3361	GGGATCATGT	AACTCGCCTT	GATCGTTGGG	AACCGGAGCT	gaatgaagcc	ataccaaacg	acgagcgtga
	CCCTAGTACA	TTGAGCGGAA	CTAGCAACCC	TTGGCCTCGA	cttacttcgg	tatggtttgc	Tgctcgcact
3431	CACCACGATG	CCTGTAGCAA GGACATCGTT	TGGCAACAAC ACCGTTGTTG	GTTGCGCAAA CAACGCGTTT	Ctattaactg Gataattgac	GCGAACTACT CGCTTGATGA	Tactctagct Atgagatcga
3501	TCCCGGCAAC	aattaataga	CTGGATGGAG	GCGGATAAAG	TTGCAGGACC	actictecec	TCGGCCCTTC
	AGGGCCGTTG	Ttaattatct	GACCTACCTC	CGCCTATTTC	AACGTCCTGG	Teaagacece	AGCCGGGAAG

SUBSTITUTE SHEET (RULE 26)

### FIG. 44F

3571	GCCGACCGAC	CGGCTGGCTG GTTTATTGCT GCCGACCGAC CAAATAACGA	gataaatctg Ctatttagac	GAGCCGGTGA	GCGTGGGTCT	CGCGGTATCA GCGCCATAGT	TTGCAGCACT AACGTCGTGA
3641	GGGGCCAGAT	GGTAAGCCCT CCATTCGGGA	CCCGTATCGT GGGCATAGCA	agttatctac Tcaatagatg	Acgacgggga Tgctgcccct	GTCAGGCAAC	Tatggatgaa Atacctactt
3711	CGAAATAGAC GCTTTATCTG	AGATCGCTGA TCTAGCGACT	gataggtgcc Ctatccacgg	TCACTGATTA AGTGACTAAT	agcattggta tcgtaaccat	actgtcagac tgacagtctg	Caagtttact Gttcaaatga
3781	Catatatact Gtatatatga	ttagattgat Aatctaacta	ttaaaacttc aattttgaag	attittaait Taaaaattaa	Taaaaggatc attttcctag	TAGGTGAAGA ATCCACTTCT	TCCTTTTTGA AGGAAAAACT
3851	Taatctcatg attagagtac	ACCAAAATCC TGGTTTTAGG	CTTAACGTGA GAATTGCACT	GTTTTCGTTC Caaaagcaag	CACTGAGCGT GTGACTCGCA	CAGACCCCGT GTCTGGGGGCA	agaaaagatc Tcttttctag
3921	AAAGGATCTT TTTCCTAGAA	CTTGAGATCC GAACTCTAGG	tittititicig Aaaaaaagac	CGCGTAATCT GCGCATTAGA	GCTGCTTGCA CGACGAACGT	aacaaaaaa Tigiiiiiiii	CCACCGCTAC
3991	CAGCGGTGGT	ttgtttgccg aacaaacggc	GATCAAGAGC CTAGTTCTCG	TACCAACTCT ATGGTTGAGA	tititccgaag aaaaggctic	GTAACTGGCT CATTGACCGA	TCAGCAGAGC AGTCGTCTCG
4061	GCAGATACCA CGTCTATGGT	AATACTGTCC TTATGACAGG	TTCTAGTGTA AAGATCACAT	GCCGTAGTTA CGGCATCAAT	GGCCACCACT	TCAAGAACTC AGTTCTTGAG	TGTAGCACCG ACATCGTGGC
4131	CCTACATACC	TCGCTCTGCT	aatcctgtta Ttaggacaat	CCAGTGGCTG GGTCACCGAC	CTGCCAGTGG GACGGTCACC	CGATAAGTCG GCTATTCAGC	tgtcttaccg acagaatggc
4201	GGTTGGACTC CCAACCTGAG	AAGACGATAG TTCTGCTATC	TTACCGGATA AATGGCCTAT	AGGCGCAGCG TCCGCGTCGC	GTCGGGCTGA ACGGGGGGTT CAGCCCGACT TGCCCCCAA	acggggggtt Tgcccccaa	CGTGCACACA GCACGTGTGT
4271	GCCCAGCTTG	GAGCGAACGA CTCGCTTGCT	CCTACACCGA	actgagatac Tgactctatg	CTACAGCGTG GATGTCGCAC	agctatgaga Tcgatactct	AAGCGCCACG TTCGCGGTGC
4341	CTTCCCGAAG GAAGGGCTTC	Ggagaaaggc Cctcttttccg	GGACAGGTAT	CCGGTAAGCG GGCCATTCGC	GCAGGGTCGG	AACAGGAGAG TTGTCCTCTC	CGCACGAGGG

FIG.\_ 44G

SUBSTITUTE SHEET (RULE 26)

GGAACAAAAG CCTTGTTTTC	CTCACTAAAG GGAACAAAAG GAGTGATTFC CCTTGTTFTC	BSSHII  ACCATGATTA CGCCAAGCGC GCAATTAACC TGGTACTAAT GCGGTTCGCG CGTTAATTGG	BESHII	accatgatta Tggtactaat	BSSHII TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAACC AGTGTGTCCT TTGTCGATAC TGGTACTAAT GCGGTTCGCG CGTTAATTGG	TCACACAGGA AGTGTGTCCT	4901
Gataacaatt Ctattgttaa	ATTGTGAGCG TAACACTCGC	GTTGTGTGGA CAACACACCT	CGGCTCGTAT	tttatgcttc aaatacgaag	AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGGA ATTGTGAGCG GATAACAATT TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTGGC CTATTGTTAA	AGGCACCCCA	4831
CTCACTCATT GAGTGAGTAA	TTCCCGACTG GAAAGCGGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TGCGTTAATT ACACTCAATC GAGTGAGTAA	acgcaattaa Tgcgttaatt	agtgagcgca Tcactcgcgt	GAAAGCGGGC CTTTCGCCCG		CACGACAGGT GTGCTGTCCA	4761
atgcagctgg tacgtcgacc	cgattcatta gctaagtaat	GCGCGTTGGC	GCCTCTCCCC	TACGCAAACC GCCTCTCCCC ATGCGTTTGG CGGAGAGGGG	ggaagcggaa gagcgcccaa tacgcaaacc gcctccccc gcgcgttggc cgattcatta atgcagctgg ccttcgcctt ctcgcgggtt atgcgttgg cggagagggg cgcgcaaccg gctaagtaat tacgtcgacc	GGAAGCGGAA CCTTCGCCTT	4691
CAGTGAGCGA GTCACTCGCT	CGCAGCGAGT GCGTCGCTCA	AACGACCGAG TTGCTGGCTC	GCCGCAGCCG	gataccectc Ctatecceae	TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA AATGGCGGAA ACTCACTCGA CTATGGCGAG CGGCGTCGGC TTGCTGGCTC GCGTCGCTCA GTCACTCGCT	ttaccgcctt Aatggcggaa	4621
gataaccgta Ctattggcat	TGATTCTGTG ACTAAGACAC	CGTTATCCCC	ttctttcctg aagaaaggac	TGCTCACATG	CTGGCCTTTT GCTGGCCTTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA GACCGGAAAA CGACCGGAAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT	CTGGCCTTTT GACCGGAAAA	4551
tttacggtec aaatgccaag	AACGCCAGCA ACGCGGCCTT TTTACGGTTC TTGCGGTCGT TGCGCCGGAA AAATGCCAAG		CCTATGGAAA GGATACCTTT	TECTCOTCAG GGGGGGGAG CCTATGGAAA ACGAGCAGTC CCCCCCCTC GGATACCTTT	ATTITITGIGA IGCTCGTCAG GGGGGCGGAG CCTAIGGAAA TAAAAACACI ACGAGCAGIC CCCCCCCTC GGAIACCTIT	atttttgtga Taaaaacact	4481
TTGAGCGTCG	CACCTCTGAC GTGGAGACTG	CGGGTTTCGC GCCCAAAGCG	ATAGTCCTGT TATCAGGACA	TGGTATCTTT ACCATAGAAA	AGCTTCCAGO GGGAAACGCC TGGTATCTTT ATAGTCCTGT CGGGTTTCGC CACCTCTGAC TTGAGCGTCG TCGAAGGTCC CCCTTTGCGG ACCATAGAAA TATCAGGACA GCCCAAAGCG GTGGAGACTG AACTCGCAGC	AGCTTCCAGG TCGAAGGTCC	4411

FIG.\_44H

EGOR CTGG GACC

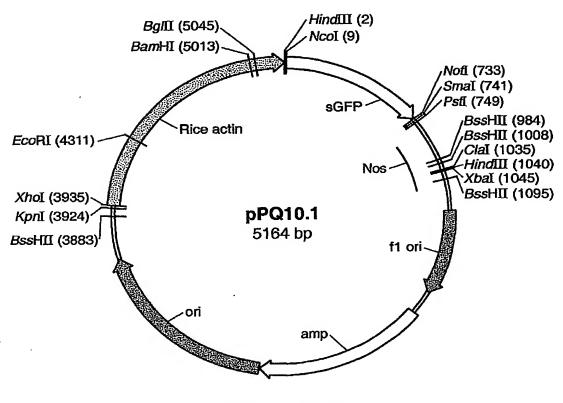


FIG.\_45A

### Sequence for pPQ10.1

u	i	n	A	T	T	т	N	_	_	T
а	_	.,	u	1	_	_	_ LY	C	c	_

1	AAGCTTACCA	TGGTGAGCAA	GGGCGAGGAG	CTGTTCACCG	GGGTGGTGCC
CATCCTGGT	C GAGCTGGAC	3			

TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG

71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC

CGCTGCACTT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA

 $141\,$  CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCACCC TCGTGACCAC CTTCACCTAC

GGACTTCAAG TAGACGTGGT GGCCGTTCGA CGGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG

211 GGCGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG

281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC AAGACCCGCG CCGAGGTGAA

TTCCGATGCA GGTCCTCGCG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG

351 GTTCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC

421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCGACAAG CAGAAGAACG

GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA CCGGCTGTTC GTCTTCTTGC

 $491\,$  GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG CAGCTCGCCG ACCACTACCA

CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCGCAC GTCGAGCGGC TGGTGATGGT .

561 GCAGAACACC CCCATCGGCG ACGGCCCCGT GCTGCTGCCC GACAACCACT ACCTGAGCAC CCAGTCCGCC

631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA

Fig. 45B

- 1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC
- 1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCGTC
- TCGCGGGATC GCGGGCGAGG AAAGCGAAAG AAGGGAAGGA AAGAGCGGTGCAAGCGGCCG AAAGGGGCAG
- 1401 AAGCTCTAAA TCGGGGGCTC CCTTTAGGGT TCCGATTTAG TGCTTTACGG CACCTCGACC CCAAAAAACT
- TTCGAGATTT AGCCCCCGAG GGAAATCCCA AGGCTAAATC ACGAAATGCC GTGGAGCTGG GGTTTTTTGA
- 1471 TGATTAGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA TAGACGGTTT TTCGCCCTTT GACGTTGGAG
- ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGGACT ATCTGCCAAA
  AAGCGGGAAA CTGCAACCTC
- 1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCCTATCTCG GTCTATTCTT
- AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGAA
- 1611 TTGATTTATA AGGGATTTTG CCGATTTCGG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTTAA
- AACTAAATAT TCCCTAAAAC GGCTAAAGCC GGATAACCAA TTTTTTACTC GACTAAATTG TTTTTAAATT
- 1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTTAGGTG GCACTTTCG
- GCGCTTAAAA TTGTTTTATA ATTGCGAATG TTAAATCCAC CGTGAAAAGC
- 1751 CTATTTGTTT ATTTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCCT GATAAATGCT
- GATAAACAAA TAAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTTACGA
- 1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCG
- AGTTÁTTATA ACTTTTTCCT TCTCATACTC ATAAGTTGTA AAGGCACAGC GGGAATAAGG GAAAAAACGC
- 1891 GCATTITGCC TTCCTGTTTT TGCTCACCCA GAAACGCTGG TGAAAGTAAA AGATGCTGAA GATCAGTTGG
- CGTAAAACGG AAGGACAAAA ACGAGTGGGT CTTTGCGACC ACTTTCATTT TCTACGACTT CTAGTCAACC

Fig. 45 C

- 1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTTC GCCCCGAAGA
- CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTCGCC ATTCTAGGAA
- 2031 ACGTTTTCCA ATGATGAGCA CTTTTAAAGT TCTGCTATGT GGCGCGGTAT TATCCCGTAT TGACGCCGGG
- TGCAAAAGGT TACTACTCGT GAAAATTTCA AGACGATACA CCGCGCCATA ATAGGGCATA ACTGCGGCCC
- 2101 CAAGAGCAAC TCGGTCGCCG CATACACTAT TCTCAGAATG ACTTGGTTGA GTACTCACCA GTCACAGAAA
- GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT
- 2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA
- TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG
- 2241 GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT
- CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GAITGGCGAA AAAACGTGTT GTACCCCCTA
- 2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA
- GTACATTGAG CGGAACTAGC AACCCTTGGC CTCGACTTAC TTCGGTATGG
- 2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA
- GCTACGGACA TCGTTACCGT TGTTGCAACG CGTTTGATAA TTGACCGCTT
- $2451\,$  GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGCGCTCGGC CCTTCCGGCT
- CGTTGTTAAT TATCTGACCT ACCTCCGCCT ATTTCAACGT CCTGGTGAAG ACGCGAGCCG GGAAGGCCGA
- 2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG
- CCGACCAAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC
  ATAGTAACGT CGTGACCCCG
- 2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA
- GTCTACCATT CGGGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCTTT
- 2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACTGT CAGACCAAGT TTACTCATAT
- ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA

Fig. 450

2731	atactittaga tatgaaatct	ttgatttaaa Aactaaattt	actecatete Tgaagtaaaa	ACTTCATTT TAATTTAAAA GGATCTAGGT TGAAGTAAAA ATTAAATTTT CCTAGATCCA	GGATCTAGGT CCTAGATCCA	GAAGATCCTT CTTCTAGGAA	TTTGATAATC AAACTATTAG
2801	TCATGACCAA AGTACTGGTT	AATCCCTTAA TTAGGGAATT	CGTGAGTTTT GCACTCAAAA	CGTTCCACTG GCAAGGTGAC		CCCGTAGAAA GGGCATCTTT	AGATCAAAGG TCTAGTTTCC
2871	atcttcttga Tagaagaact	Gatccttttt Ctaggaaaa	TTCTGCGCGT AAGACGCGCA	AATCTGCTGC TTGCAAACAA TTAGACGACG AACGTTTGTT	ttgcaaacaa aacgtttgtt	AAAAACCACC TTTTTGGTGG	GCTACCAGCG CGATGGTCGC
2941	GIGGITITGIT CACCAAACAA	GTGGTTTGTT TGCCGGATCA CACCAAACAA ACGGCCTAGT		actctttttc Tgagaaaaag	CGAAGGTAAC GCTTCCATTG	TGGCTTCAGC ACCGAAGTCG	AGAGCGCAGA TCTCGCGTCT
3011	Taccaaatac Atgetttate	TGTCCTTCTA	GTGTAGCCGT CACATCGGCA	AGTTAGGCCA CCACTTCAAG TCAATCCGGT GGTGAAGTTC	CCACTTCAAG GGTGAAGTTC	AACTCTGTAG TTGAGACATC	Caccectac Gegecegate
3081	atacctcgct tatggagcga	CTGCTAATCC GACGATTAGG	TGTTACCAGT ACAATGGTCA	TGTTACCAGT GGCTGCTGCC AGTGGCGATA ACAATGGTCA CCGACGACGG TCACCGCTAT	agtggcgata Tcaccgctat	AGTCGTGTCT TCAGCACAGA	Taccgggttg Atggcccaac
3151	GACTCAAGAC CTGAGITCTG	Gatagitacc Ctatcaatgg		GGATAAGGCG CAGCGGTCGG GCTGAACGGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC		GGGTTCGTGC CCCAAGCACG	ACACAGCCCA TGTGTCGGGT
3221	GCTTGGAGCG CGAACCTCGC	AACGACCTAC TTGCTGGATG	AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT	GATACCTACA CTATGGATGT		TGAGAAAGCG ACTCTTTCGC	CCACGCTTCC
3291	CGAAGGGAGA	AAGGCGGACA TTCCGCCTGT	GGTATCCGGT CCATAGGCCA	GGTATCCGGT AAGCGGCAGG GTCGGAACAG CCATAGGCCA TTCGCCGTCC CAGCCTTGTC		GAGAGCGCAC CTCTCGCGTG	GAGGGAGCTT CTCCCTCGAA
3361	CCAGGGGAA GGTCCCCCTT	ACGCCTGGTA TGCGGACCAT	tctttatagt agaaatatca	TCTTTATAGT CCTGTCGGGT TTCGCCACCT AGAAATATCA GGACAGCCCA AAGCGGTGGA	TTCGCCACCT AAGCGGTGGA	CTGACTTGAG GACTGAACTC	CGTCGATTTT GCAGCTAAAA
3431	TGTGATGCTC ACACTACGAG		GTCAGGGGGG CGGAGCCTAT GGAAAACGC CAGCAACGCG GCCTTTTAC CAGTCCCCCC GCCTCGGATA CCTTTTTGCG GTCGTTGCGC CGGAAAAATG	GGAAAAACGC CCTTTTTGCG	CAGCAACGCG GTCGTTGCGC		GGTTCCTGGC

## FIG.\_ 45E

3501	CTTTTGCTGG GAAACGACC	CCTTTTGCTC GGAAAACGAG	CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA CCGTATTACC GAAAACGACC GGAAAACGAG TGTACAAGAA AGGACGAAT AGGGGACTAA GACACCTATT GGCATAATGG	TCCTGCGTTA AGGACGCAAT	TCCCCTGATT AGGGGACTAA	TCCCCTGATT CTGTGGATAA AGGGGACTAA GACACCTATT	CCGTATTACC GGCATAATGG
3571	GCCTTTGAGT CGGAAACTCA	gagctgatac Ctcgactatg	GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGCTCGCGTC GCTCAGTCAC TCGCTCCTTC	agccgaacga Tcggcttgct	CCGAGCGCAG GGCTCGCGTC	CGAGTCAGTG GCTCAGTCAC	agcgaggaag Tcgctccttc
3641	CGGAAGAGCG GCCTTCTCGC		CCCAATACGC AAACCGCCTC GGGTTATGCG TTTGGCGGAG	TCCCCGCGCG	TTGGCCGATT AACCGGCTAA	CATTAATGCA GCTGGCACGA GTAATTACGT CGACCGTGCT	GCTGGCACGA
3711	CAGGTTTCCC GTCCAAAGGG		CGGGCAGTGA GCCCGTCACT	GCGCAACGCA CGCGTTGCGT	attaatgtga Taattacact	GTTAGCTCAC CAATCGAGTG	TCATTAGGCA AGTAATCCGT
3781	CCCCAGGCIT	tacactttat Atgegaaata	GCTTCCGGCT	CGTATGTTGT GCATACAACA	GTGGAATTGT CACCTTAACA	GAGCGGATAA CTCGCCTATT	Caatttcaca Gttaaagtgt
3851	CAGGAAACAG GTCCTTTGTC	CAGGAAACAG CTATGACCAT GTCCTTTGTC GATACTGGTA	GATTACGCCA CTAATGCGGT	Bashii ~~~~~ AGCGCGCAAT TCGCGCGTIA	TAACCCTCAC ATTGGGAGTG	Kpri  Trargggrac raragetegg Atttcceteg ttttcgrcc	Kpri aaaagctggg Tfttcgaccc
3921	Kpri Trccegecc Afeccege		TCATTCATAT AGTAAGTATA	gcttgagaag cgaactcttc	agagtcggga TCTCAGCCCT	GCTTGAGAAG AGAGTCGGGA TAGTCCAAAA TAAAACAAAG CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTTGTTTC	Taracarg Ativitgivic
3991	GTAAGATTAC CATTCTAATG	CTGGTCAAAA GACCAGTTTT	CTGGTCAAAA GTGAAAACAT GACCAGTTTT CACTTTTGTA	Cagttaaaag Gtcaattttc	GTGGTATAAG CACCATATTC	TAAAATATCG GTAATAAAAG ATTTTATAGC CATTATTTC	GTAATAAAG CATTATTTTC
4061	GTGGCCCAAA CACCGGGTTT	GTGAAATTTA CACTTTAAAT	GTGAAATTTA CTCTTTTCTA CTATTATAAA AATTGAGGAT CACTTTAAAT GAGAAAGAT GATAATATTT TTAACTCCTA	Ctattataaa Gataatatt	aattgaggat Ttaactccta	GTTTTGTCGG TACTTTGATA CAAAACAGCC ATGAAACTAT	tactttgata Atgaaactat

## -1G.\_45F

4131	CGTCATTTT GCAGTAAAAA		GTATGAATTG GTTTTTAAGT TTATTCGCGA CATACTTAAC CAAAAATTCA AATAAGCGCT	<b>TTATTCGCGA</b> AATAAGCGCT	tttggaaatg Aaacctttac	CATATCTGTA GTATAGACAT	tttgagtcgg aaactcagcc
4201	tttttaagtt Aaaaattcaa	CGTTCCTTT GCAACGAAAA	GTAAATACAG AGGGATTTGT CATTTATGTC TCCCTAAACA		ataagaaata Tattctttat	tctttaaaa Agaaatttit	acccatatec tegetatace
4271	taatttgaca Attaaactge	taattttyga Attaaaaact	gaaaaatata Ctittibibi	EGORI TATTCAGGCG AATTCCACAA ATAAGTCCGC TTAAGGTGTT	EGORI G AATTCCACAA C TTAAGGTGTT	Tgaacaataa Acttgetatt	taagattaaa Aftctaaitt
4341	ATAGCTTGCC TATCGAACGG		CCCGTTGCAG CGATGGGTAT GGGCAACGTC GCTACCCATA	tttttctagt aaaaagatca	aaaataaag Ttttatititc	ataaacttag Tatttgaatc	actcaaaaca tgagttttgt
4411	TTTACAAAAA AAATGTTTT CCCAACCCAA GGGTTGGGTT	CAACCCTAA GTTGGGGATT CCCAACCCAC GGGTTGGGTG	AGTCCTAAAG TCAGGATTTC CCCAGTGCAG GGGTCACGTC	CCCAAAGTGC GGGTTTCACG CCAACTGGCA GGTTGACCGT	TATGCACGAT ATACGTGCTA AATAGTCTCC TTATCAGAGG	CCATAGCAAG GGTATCGTTC ACCCCGGCA TGGGGGCCGT	CCCAGCCCAA GGGTCGGGTT CTATCACCGT GATAGTGGCA
4551	GAGTTGTCCG CTCAACAGGC		CACCACCGCA CGTCTCGCAG GTGGTGGCGT GCAGAGCGTC		CCARABARA AAAAAGAAAG AAAAAAAAA GGITITITITI IIITITICITIC IITITITITITICI	aaaaaaaga TTTTTTTTTT	aaaagaaaaa Titicititi
4621	CAGCAGGTGG GTCGTCCACC		GTCCGGGTCG TGGGGGCCGG AAAAGCGAGG AGGATCGCGA CAGGCCCAGC ACCCCGGCC TTTTCGCTCC TCCTAGCGCT	AAAAGCGAGG TTTTCGCTCC	AGGATCGCGA TCCTAGCGCT	GCAGCGACGA	<u> </u>
4691	rcccrccct aggeaggcea	TCCAAAGAAA AGGTTTCTTT	CGCCCCCAT	CGCCACTATA TACATACCCC GCGGTGATAT ATGTATGGGG	TACATACCCC	CCCCTCTCCT	CCCATCCCCC
4761	CAACCCTACC	CAACCCTACC ACCACCACCA CCACCACCTC GTTGGGATGG TGGTGGTGGT GGTGGTGGTGGAG	CCACCACCTC	CTCCCCCTC GCTGCCGGACGAACGAAGGGGGGGAAG	GCTGCCGGAC	GACGAGCTCC CTGCTCGAGG	TCCCCCCTCC
4831	CCCTCCGCCG	CCGCCGGTAA GGCGGCCATT	CCACCCCGCC		CCTCTCCTCT TTCTTTCTCC GGAGAGAAAAAAGG	Gptytytyty Caaaaaaaa	<b>TCGTCTCGGT</b> AGCAGAGCCA

-i -> 	GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC	CCGGAACCAT	CAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC	CGCTCTCGCC	GAAGCAGCGG	GTCTAGCCAC	CGCCCTCCC	
					BenHI			
4971	4971 GCGGGARCTC GCGGCTGGCG TCTCCCGGCC TGAGTCGCCC CGGATCCTCG CGGGGAATGG GCCTCTCGGA	GCGGCTGGCG	TCTCCGGGCG	TGAGTCGGCC	CGGATCCTCG	CGGGGAATGG	GGCTCTCGGA	
	CGCCCTAGAG	CGCCGACCGC	COCCCTAGAG COCCGACCGC AGAGGCCCGC ACTCAGCCGG GCCTAGGAGC GCCCCTTACC CCGAGAGCCT	ACTCAGCCGG	GCCTAGGAGC	GCCCCTTACC	CCGAGAGCCT	
	Bglii							
	1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2							

TGTAGATCTT CTTTCTTTCT TCTTTTTGTG GTAGAATTTG AATCCCTCAG CATTGTTCAT CGGTAGTTTT ACATCTAGAA GAAAGAAAGA AGAAAACAC CATCTTAAAC TTAGGGAGTC GTAACAAGTA GCCATCAAAA

5041

5111

TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTGCGGAGC TTTTTTGTAG GTAG AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCTCG AAAAAACATC CATC

FIG.\_45H

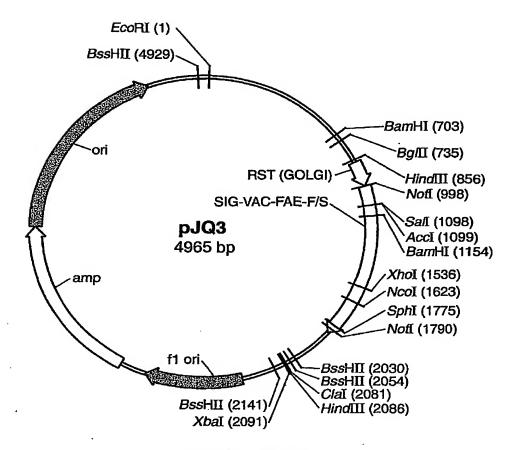


FIG.\_46A

CGATGGGTAT TTTTTCTAGT GCTACCCATA AAAAAGATCA	AGTCCTAAAG CCCAAAGTGC TCAGGATTTC GGGTTTCACG	CCCAGTGCAG CCAACTGGCA GGGTCACGTC GGTTGACCGT	CGTCTCGCAG CCAAAAAAA GCAGAGCGTC GGTTTTTTTTT	TGGGGGCGG AAAAGCGAGG ACCCCGGGCC TTTTCGCTCC	CCCCCCAT CCCCACTATA GCCCCGGGGGTA GCCGTGATAT	CCACCACCTC CTCCCCCTC GGTGGTGGAG GAGGGGGGAG	CCACCCGCC CCTCTCCTCT GGTGGGGCGG GGAGGAGGAGA	GTTTGGGTGG GCGAGAGCGG CAAACCCACC CGCTCTCGCC	TCTCCGGGCG TGAGTCGGCGG	TCTTTTTGTG GTAGAATTTG AGAAAAACAC CATCTTAAAC
CCCGTTGCAG CC GGGCAACGTC GC	CAACCCCTAA ACGTTGGGGGATT TC	CCCAACCCAC GGGTTGGGTG	CACCACCGCA GTGGTGGCGT	GTCCGGGTCG	TCCCTCCGCT TCCAAAGAAA CC AGGGAGGCGA AGGTTTCTTT GC	CAACCCTACC ACCACCACCA GTTGGGATGG TGGTGGTGGT G	CCGCCGGTAA	GGCCTTGGTA	GCGGCTGGCG TY	CTTTCTTTCT T
atagcttgcc tatcgaacgg	tttacaaaaa Aaatgitttt	CCCAGCCCAA CCCAACCCAA GGGTCGGGTT GGGTTGGGTT	CTCAACAGGC	CAGCAGGTGG GTCGTCCACC	TCCCTCCGCT AGGGAGGCGA			CTCGATCTTT GAGCTAGAAA	GCGGGGATCTC	BGLII TGTAGATCTT ACATCTAGAA
taagattaaa Attctaattt	actcaaaaca Tgagttttgt	CCCAGCCCAA GGGTCGGGTT	CTATCACCGT GATAGTGGCA	aaaagaaaa Ttttctttt	ອອອວວອອອວວ ວວວອອວວວອອ	CCCATCCCCC	TCCCCCCTCC	TCGTCTCGGT	CGCGGGAGGG	GGCTCTCGGA CCGAGAGCCT
TGAACAATAA ACTTGTTATT	ATAAACTTAG TATTTGAATC	CCATAGCAAG GGTATCGTTC	ACCCCCGGCA	aaaaaaaga Tititititee	GCAGCGACGA	CCCCTCTCCT	GACGAGCTCC	Gttttttttt Caaaaaaaa	Cagatcegte Gtctagccac	CGGGGAATGG GCCCTTACC
AATTCCACAA TTAAGGTGTT	AAAATAAAG . TTTTATTTTC	TATGCACGAT ATACGTGCTA	AATAGTCTCC TTATTCAGAGG	AAAAAGAAG TTTTCTTTC	AGGATCGCGA	TACATACCCC	GCTGCCGGAC	TTCTTTCTCC	CTTCGTCGCC	Bamki cccarcctcc
Ħ	71	141	211	281	351	421	491	561	631	701

771 AATCCCTCAG CATTGTTCAT CGGTAGTTTT TCTTTTCATG ATTGTGACA AATGCAGCCT CGTGCGGAGC TTAGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCTCG	
aatgcagcct Ttacgtcgga	
atttgtgaca Taaacactgt	
tcitticatg Agaaaagtac	
CGGTAGTTTT GCCATCAAAA	
CATTGTTCAT GTAACAAGTA	Hindrin
aatccctcag Ttagggagtc	
771	

CAAAAAAAA TTCTCCCTCT TCATCCTCGT GTTTTTCTTC AAGAGGGAGA AGTAGGAGCA GTAGAAGCTT ACCATGATCC ACACCAACCT TGTGGTTGGA TGGTACTAGG CATCTTCGAA TTTTTGTAG AAAAAACATC 841

CTTCCTCCTC TTCGCCGTGA TCTGCGTGTG GAAGAAGGGC TCCGACTACG AGGCCCTCAC CCTCCAAGCC GAAGGAGGAG AAGCGGCACT AGACGCACAC CTTCTTCCCG AGGCTGATGC TCCGGGAGTG GGAGGTTCGG 911

Noti

\*\*\*\*\*\*\*\*\*

981

AAGGAGTTCC AAATGGCGGC CGCCTCCACG CAGGGCATCT CCGAAGACCT CTACAGCCGT TTAGTCGAAA TTCCTCAAGG TTTACCGCCG GCGGAGGTGC GTCCCGTAGA GGCTTCTGGA GATGTCGGCA AATCAGCTTT

222222 SalI

Acci

OCCIACGCCG ACCIGIGCAN CATICCGICG ACTALITATCA AGGGAGAGAA GTAAGGCAGC TGATAATAGT TCCCTCTCTT TGGCCACTAT CTCCCAAGCT GCCTACGCCG ACCTGTGCAAAACCGGTGATA GAGGGTTCGA CGGATGCGGC TGGACACGTT

BankI

CGCGACGACA GCAGCAAAGA AATAATCACC GCGCTGCTGT CGTCGTTTCT TTATTAGTGG AATTTACAAT TCTCAAACTG ACATTAACGG ATGGATCCTC TTAAATGTTA AGAGTTTGAC TGTAATTGCC TACCTAGGAG 1121

GTGGGAGTGC GGAAAGCTGT CCTTTCGACA CACCCTCACG ATACTAACTA COTGACCATC ACTATGCTTA GATGTTGAGC TATGATTGAT CTACAACTCG TGATACGAAT GCACTGGTAG GTCTTCCGTG CAGAAGGCAC 1191

TGGGTCTCCG TCCAGGACCA TACGTTGCCA ACACTTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCCTGGT CCCTACCACA ATGCAACGGT TGTGAAGTAC ACGGTGGATA TTATATTGGA GGGATGGTGT 1261

CTTGTCAAAC AGCAGGTTAG CCAGTATCCG GAACAGTTTG TCGTCCAATC GGTCATAGGC CCCTGGCGGC ACTCACTGCC GCCCAGCTGT GGGACCGCCG TGAGTGACGG CGGGTCGACA
ACTCACTGCC TGAGTGACGO GGCAATCAGG
CCGTTAGTCC
ACCCACATT TCCGGGTCAC TCATGCCAC TGCGTCATAA AGGCCCAGTG AGTACGGTTG
NGOI CCCATGGCGG TGTAGAGTAC TGGAGCGTTG GGGTACCGCC ACATCTCATG ACCTCGCAAC
TGAAGTGCAG TGCTGTGAGG CCCAGGGCGG ACAGGGTGTG AATAATGCGC ACTTCACGTC ACGACACTCC GGGTCCCGCC TGTCCCACAC TTATTACGCG
Sphi
ACCTG GCCGGTC
TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AGCAAGTTTG TAAACCGTTA TYTCAAAGAA
TAATTTCTGT TGAATTACGT TAAGCATGTA ATTAAAGACA ACTTAATGCA ATTCGTACAT

# FIG.\_46D

							TTUGER
1961	ATGAGATGGG TACTCTACCC		TTTTATGAT TAGAGTCCCG CAATTATACA AAAAATACTA ATCTCAGGGC GTTAATATGT	CAATTATACA GTTAATATGT	tttaatacgc aaattatgcg	TTTAATACGC GATAGAAAAC AAATTATGCG CTATCTTTG	AAAATATAGC TTTTATATCG
							Xbal
	Bashii		BSSHII		ទ	clar Hindiri	zzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzzz
2031	GCGCAAACTA	GGATAAATTA CCTATTTAAT	TCGCGCGCGG	TGTCATCTAT GTTACTAGAT CGATAAGCTT ACAGTAGATA CAATGATCTA GCTATTCGAA	GTTACTAGAT CAATGATCTA	TGTCATCTAT GTTACTAGAT CGATAAGCTT ACAGTAGATA CAATGATCTA GCTATTCGAA	Ctrgagege Gatetegeeg
				BB	BSSKII		
2101	CGGTGGAGCT GCCACCTCGA		CCAATTCGCC CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC GGTTAAGCGG GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG	CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA	G CGCGCTCACT C GCGCGAGTGA	GGCCGTCGTT CCGGCAGCAA	TTACAACGTC AATGTTGCAG
2171	GTGACTGGGA	AAACCCTGGC TTTGGGACCG	AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TTTGGGAACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA GGGGGAAAGC GGTCGACGGC	GTTACCCAAC TTAATCGCCT CAATGGGTTG AATTAGCGGA	TGCAGCACAT CCCCCTTTCG ACGTCGTGTA GGGGGAAAGC	CCCCTTTCG	CCAGCTGGCG
2241	Taatagcgaa Attatcgctt	GAGGCCCGCA CTCCGGGCGT	TAATAGCGAA GAGGCCGGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG ATTATCGCTT CTCCGGGCGT GGCTAGCGGG AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCCTGCGC	CCGATCGCCC TYCCCAACAGGGCTAGTC	TTGCGCAGCC TGAATGGCGA AACGCGTCGG ACTTACCGCT	TGAATGGCGA ACTTACCGCT	ATGGGACGCG TACCCTGCGC
2311	CCCTGTAGCG		GCGCATTAAG CGCGGCGGGT GTGGTGGTTA CGCGTAATTC GCGCCGCCCA CACCACAAT	GTGGTGGTTA CACCACCAAT	CGCGCAGCGT	CGCGCAGCGT GACCGCTACA CTTGCCAGCG GCGCGTCGCA CTGGCGATGT GAACGGTCGC	CTTGCCAGCG GAACGGTCGC
2381	CCCTAGCGCC	CGCTCCTTTC		GCTTTCTTCC CTTCCTTTCT CGAAAGAAGG GAAGGAAAGA	CGCCACGTTC GCCGGCTTTC GCGGTGCAAG CGGCCGAAAG	GCCGGCTTTC	CCCGTCAAGC GGGCAGTTCG
2451	TCTRARTCGG AGATTTRGCC	GGGCTCCCTT		TAGGGTTCCG ATTTAGTGCT ATCCCAAGGC TAAATCACGA	TTACGGCACC TCGACCCCAA AATGCCGTGG AGCTGGGGTT		aaacttgat Ttttgaacta
2521	TAGGGTGATG ATCCCACTAC	GTTCACGTAG CAAGTGCATC	GITCACGIAG IGGGCCATCG CCCIGATAGA CGGITTITICG CCCITIGACG ITGGAGICCA CAAGIGCAIC ACCCGGIAGC GGGACIAICI GCCAAAAAGC GGGAAACIGC AACCICAGGI	CCCTGATAGA	CGGTTTTTCG CCCTTTCACG GCCAAAAACC GGGAAACTGC	CCCTTTGACG GGGAAACTGC	TTGGAGTCCA AACCTCAGGT

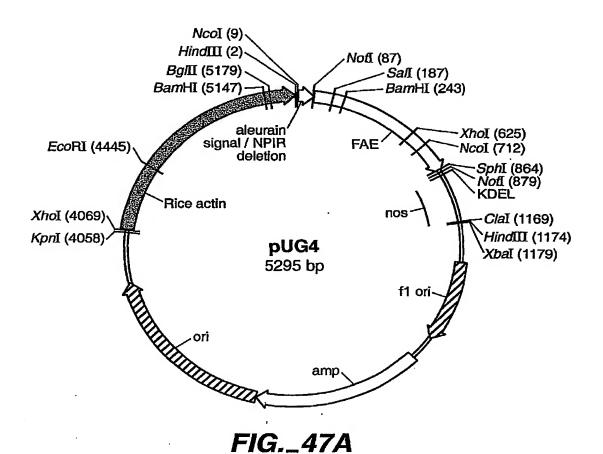
### FIG. 46E

ACACCACGAT TGTGGTGCTA	GACGAGCGTG	Cataccaaac Gtatggtttg	TGAATGAAGC ACTTACTTCG	GAACCGGAGC	TGATCGTTGG	TAACTCGCCT	3361
GGGGATCATG CCCCTAGTAC	GCACAACATG CGTGTTGTAC	CCGCTTTTTT GGCGAAAAAA	aaggagctaa Ttcctcgatt	CGGAGGACCG GCCTCCTGGC	TGACAACGAT ACTGTTGCTA	aactitactic Tigaatgaag	3291
CACTGCGGCC	tgagtgataa Actcactatt	GCCATAACCA CGGTATTGGT	atgcagtgct tacgtcacga	taagagaatt attetettaa	GGCATGACAG CCGTACTGTC	TCTTACGGAT AGAATGCCTA	3221
CAGAAAAGCA GTCTTTTCGT	<b>TCACCAGTCA</b> AGTGGTCAGT	GGTTGAGTAC CCAACTCATG	AGAATGACTT TCTTACTGAA	Cactapticic Gigataagag	TCGCCGCATA AGCGGCGTAT	AGCAACTCGG TCGTTGAGCC	3151
GCCGGCCAAG	CCGTATTGAC GCCATAACTG	CGGTATTATC GCCATAATAG	Ctatgtgggg Gatacaccgc	TAAAGTTCTG ATTTCAAGAC	TGAGCACTTT ACTCGTGAAA	TITCCAATGA AAAGGTTACT	3081
CGAAGAACGT GCTTCTTGCA	GTTTTCGCCC CAAAAGCGGG	atccttgaga taggaactct	CAGCGGTAAG GTCGCCATTC	TGGATCTCAA ACCTAGAGTT	Tacatcgaac Atgeagcetg	ACGAGTGGGT	3011
agtigggigc Tcaacccacg	GCTGAAGATC	agtaaaagat Tcattttcta	CGCTGGTGAA GCGACCACTT	CACCCAGAAA GTGGGTCTTT	TGTTTTTGCT ACAAAAACGA	TTTGCCTTCC AAACGGAAGG	2941
tttgcggcat Aaacgccgta	TATTCCCTTT ATAAGGGAAA	GTGTCGCCCT	CAACATTTCC GTTGTAAAGG	tatgagtatt Atactcataa	aaaggaagag Titccttctc	Taatattgaa Attataactt	2871
aatgcttcaa ttacgaagtt	AACCCTGATA TTGGGACTAT	atgagacaat tactctgtta	Gtatcccctc Cataggcgag	attcaaatat Taagtttata	tectaratac Aagatetaeg	ttgittatt aacaaataaa	2801
GAACCCCTAT CTTGGGGATA	aatgtgcgcg Ttacacgcgc	TTTTCGGGGA AAAAGCCCCT	TAGGTGGCAC ATCCACCGTG	GCTTACAATT CGAATGTTAA	aaatattaac Tttataattg	aatititaaca Itaaaatigi	2731
atttaacgcg taaattgcgc	tttaacaaaa Aaattgittt	aatgagctga Ttactcgact	ttggttaaaa aaccaatttt	TTTCGGCCTA AAAGCCGGAT	attttgccga taaaacggct	TTTATAAGGG AAATATTCCC	2661
attcttttga taagaaaact	ATCTCGGTCT TAGAGCCAGA	ACTCAACCCT TGAGTTGGGA	CTGGAACAAC GACCTTGTTG	TTGTTCCAAA AACAAGGTTT	TAGTGGACTC ATCACCTGAG	CGTTCTTTAA GCAAGAAATT	2591

3431	GCCTGTAGCA	Atgccaacaa Taccgttgtt	CGTTGCGCAA GCAACGCGTT	actattaact Tgataattga	GGCGAACTAC	TTACTCTAGC AATGAGATCG	TTCCCGGCAA AAGGGCCGTT
3501	Caattraatag Gitaattaec	ACTGGATGGA TGACCTACCT	gecggataaa Ccccctattt	GTTGCAGGAC CAACGTCCTG	CACTTCTGCG GTGAAGACGC	CTCGGCCCTT	CCGGCTGGCT
3571	ggittattgc Ccaaataacg	tgataaatct actatttaga	GGAGCCGGTG	AGCGTGGGTC TCGCACCCAG	TCGCGGTATC AGCGCCATAG	ATTGCAGCAC TAACGTCGTG	TGGGGCCAGA
3641	TGGTAAGCCC ACCATTCGGG	TCCCGTATCG AGGGCATAGC	tagitatcta Atcaatagat	CACGACGGGG	agtcaggcaa Tcagtccgtt	Ctatggatga Gatacctact	acgaaataga tgctttatct
3711	CAGATCGCTG GTCTAGCGAC	agataggtgc Tctatccacg	CTCACTGATT GAGTGACTAA	aagcattggt ttcgtaacca	aactgtcaga Ttgacagtct	CCAAGTTTAC GGTTCAAATG	tcatatatac Agtatatatg
3781	tttagattga Aaatctaact	tetaaaacte Aaatetegaa	Cat'it'it'iaat Gtaaaaatta	ttaaaaggat aattttccta	Ctaggtgaag Gatccacttc	atcctttttg taggaaaaac	ataatctcat Tattagagta
3851	GACCAAAATC CTGGTTTTAG	CCTTAACGTG GGAATTGCAC	agititicgit Icaaaagcaa	CCACTGAGCG GGTGACTCGC	TCAGACCCCG	tagaaaagat Atctttttcta	Caaaggatct Gtttcctaga
3921	TCTTGAGATC AGAACTCTAG	Ctttttttct Gaaaaaaaga	GCGCGTAATC CGCGCATTAG	TGCTGCTTGC ACGACGAACG	aaacaaaaa Tttgttttt	ACCACCGCTA TGGTGGCGAT	CCAGCGGTGG
3991	TTTGTTTGCC AAACAAACGG	GGATCAAGAG CCTAGTTCTC	CTACCAACTC GATGGTTGAG	tttttccgaa Aaaaaggctt	GGTAACTGGC CCATTGACCG	TTCAGCAGAG AAGTCGTCTC	CGCAGATACC GCGTCTATGG
4061	aaatactgtc tttatgacag	CTTCTAGTGT	agccgtagtt Tcggcatcaa	AGGCCACCAC TCCGGTGGTG	ttcaagaact Aagttcttga	CTGTAGCACC GACATCGTGG	GCCTACATAC CGGATGTATG
4131	CTCGCTCTGC GAGCGAGACG	taatcctgft attaggacaa	ACCAGTGGCT TGGTCACCGA	GCTGCCAGTG CGACGGTCAC	GCGATAAGTC CGCTATTCAG	GTGTCTTACC CACAGAATGG	GGGTTGGACT CCCAACCTGA
4201	Caagacgata Gttctgctat	GTTACCGGAT CAATGGCCTA	AAGGCGCAGC TTCCGCGTCG	GGTCGGGCTG CCAGCCCGAC	AACGGGGGGT TTGCCCCCCA	AACGGGGGGT TCGTGCACAC TTGCCCCCCA AGCACGTGTG	AGCCCAGCTT TCGGGTCGAA

4271	GGAGCGAACG	ACCTACACCG TGGATGTGGC	GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTCGCA CTCGATACTC TTTCGCGGTG CGAAGGGCTT	CCTACAGCGT GGATGTCGCA	GAGCTATGAG CTCGATACTC	AAAGCGCCAC GCTTCCCGAA TYTCGCGGTG CGAAGGGCTT	GCTTCCCGAA CGAAGGGCTT
4341	GGGAGAAAGG CCCTCTTTCC	CGGACAGGTA GCCTGTCCAT	GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG CCCTCTTTCC GCCTGTCCAT AGGCCATTCG CCGTCCCAGC CTTGTCCTCT CGCGTGCTCC CTCGAAGGTC	GGCAGGGTCG	GAACAGGAGA CTTGTCCTCT	GCGCACGAGG	GAGCTTCCAG CTCGAAGGTC
4411		CTGGTATCTT GACCATAGAA	GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG CCCCTTTGCG GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTCGCAG CTAAAAACAC	TCGGGTTTCG AGCCCAAAGC	CCACCTCTGA GGTGGAGACT	CTTGAGCGTC GAACTCGCAG	gattititgeg Ctaaaaacac
4481	ATGCTCGTCA TACGAGCAGT	GGGGGGCGGA	GGGGGGGGA GCCTATGGAA AAACGCCAGC AACGCGGCT TTTTACGGTT CCTGGCCTTT CCCCCCGCT CGGATACCTT TTTGCGGTCG TTGCGCCGGA AAAATGCCAA GGACCGGAAA	AAACGCCAGC TTTGCGGTCG	AACGCGGCCT TTTTACGGTT CCTGGCCTTT TTGCGCCGGA AAAATGCCAA GGACCGGAAA	tititacggit Aaaatgccaa	CCTGGCCTTT GGACCGGAAA
4551	TGCTGGCCTT	TTGCTCACAT BACGAGTGTA	tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcct acgaccggaa aacgagtgta caagaaagga cgcaataggg gactaagaca cctattggca taatggcgga	GCGTTATCCC CGCAATAGGG	CTGATTCTGT GACTAAGACA	ggataaccot cctattogca	attaccecct taatgecgea
4621	TTGAGTGAGC AACTCACTCG	Tgataccect actategcea	TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGAA AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTGCTGGCT CGCGTCGCTC AGTCACTCGC TCCTTCGCCT	GAACGACCGA CTTGCTGGCT	GCGCAGCGAG CGCGTCGCTC	TCAGTGAGCG AGGAAGCGGA AGTCACTCGC TCCTTCGCCT	aggaagcgga Tccttcgcct
4691	AGAGCGCCCA TCTCGCGGGT	atacgcaaac tatgcgtttg	ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCAGCTG GCACGACAGG TATGCGTTTG GCGGAGAGGG GCGCGAACC GGCTAAGTAA TTACGTCGAC CGTGCTGTCC	CGCGCGTTGG CCGATTCATT GCGCGCAACC GGCTAAGTAA	CCGATTCATT GGCTAAGTAA	aatgcagctg ttacgtcgac	OCACGACAGG CGTGCTGTCC
4761	TTTCCCGACT AAAGGGCTGA	GGRAAGCGGG CCTTTCGCCC	TTTCCCGACT GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AAAGGGCTGA CCTTTCGCCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA ATCCGTGGG	aacgcaatta Ttgcgttaat	atgtgagtta tacactcaat	GCTCACTCAT	TAGGCACCCC
4831	aggctttaca Tccgaaatgt	Ctttatgctt Gaaatacgaa	aggetttaca etttatgett ecggetegta tgitgtgtgg aattgigage ggataacaat iteacaegg Teegaaatgt gaaataegaa ggeegageat acaacacaee ttaacaeteg eetattgita aagtgigtee	TGTTGTGTGG ACAACACACC	AATTGTGAGC GGATAACAAT TTCACACAGG TTAACACTCG CCTATTGTTA AAGTGTGTCC	ggataacaat Cctattgtta	TTCACACAGG AAGTGTGTCC
			BESHII	ııı			ECORI
4901	AAACAGCTAT TTTGTCGATA	GACCATGATT CTGGTACTAA	AAACAGCTAT GACCATGATT ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAACAAAA GCTGG TTTGTCGATA CTGGTACTAA TGCGGTTCGC GCGTTAATTG GGAGTGATTT CCCTTGTTTT CAACT	CGCAATTAAC	CCTCACTAAA	GGGAACAAAA CCCTTTGTTTTT	GCTGG

# FIG.\_46H



SUBSTITUTE SHEET (RULE 26)

NGOI

HindIII

TGGCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCC ø 4 ы H > p; NotI 4 AAGCTTACCA

GCCTCCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT W þ SalI r o Œ ĽΩ U Ø ರಿದಿದ್ದರಿದ್ದರಿದ್ದರ A A TCGCCTCCTC Ŋ ₫. 77

G E K • A T I S Q A A Y A D L C N I P S T I I K GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA BamHI 141

\*\*\*\*\*\*

I Y N S Q T D I N G W I L R D D S S K E I I T Y ATTIACAATT CTCAAAGAA ATAATCACCG ICTICCOTGG CACTGGTAGT GATACGAATC TACAACTCGA TACTAACTAC ACCCTCACGC CTTTCGACA O 211 281

CCTACCACAA TGCAACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA 四四 Ø > 0

TOGGOGOCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TGCGACATAC GACAACATCC GCCTGTACAC GICGAGICGC ITGICAAACA GCAGGITAGC CAGTATCCGG ACTACGCGCT GACCGIGACC GGCCACKCCC O. 421 491

Ø · F G E P R S G N Q A F A S Y M N D A F Q CTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTCGCGTCG TACATGAACG ATGCCTTCCA ĸ p, Ω Ε 561

PDTT QYFF RVTHANDGIP NLPPVEC 631

22222

FIG.\_47B

351

CGGGCAAGAG

TTTCGCCCCG

CCTTGAGAGT GTATTATCCC TTGAGTACTC CATAACCATG CCTTTTTTCC

> ATGTGGCGCG AATGACTTGG

GCGGTAAGAT

TAAAAGATGC

GTATTGACGC

GAAAAGCATC CTGCGGCCAA GGATCATGTA ACCACGATGO CCCGGCAACA

> CGAGCGTGAC ACTCTAGCTT

TACCAAACGA

ACAACATGGG AGTGATAACA ACCAGTCACA

TGCCAGCGCC

AACTTGATTA GGAGITCCACG TCTTTTGALT

GACCCCAAAA

CTTTGACGTT CTCGGTCTAT

CGGCTTTCCC

CCACGTTCGC

ACGGCACCTC GITTITICCCC TCAACCCTAT

CGTCAAGCTC

ACCCCTATT

TGCTTCAATA TGCGGCATT TTGGGTGCAC AAGAACGTTT

CCCTGATAAA TTCCCTTTTT TGAAGATCAG

GICGCCCLTA

TTAACGCGAA

TAACAAAAT TGTGCGCGGA

TGAGCTGATT

GGAACAACAC GGTTAAAAA

TCGGCCTATT

TTACAATTTA

CTGATAGACG

TTCGGGGAAA GAGACAATAA

> ATCCGCTCAT ACATTTCCGT CTGGTGAAAG

TCAAATATGT

CCCAGAAACG

GATCTCAACA AAGTTCTGCT

TGAGTATTCA

GGTGGCACTT

GCCCAGAACA CATTTGTCTG CACGACTTAT CTTGCGATGA CGTTATTTAT GAGCTGTAAA П H 闰 Ħ TITGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTGGGG CCGCGGAACC ACTGAAGGAT CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCGCA TTATCATATA ATTTCTGTTG AATTACGTTA AGCATGTAAT AATTAACATG TAATGCATGA TGTTGCCGGT A Z N H H O ď Z TCCTTACAGC GTICAAACAI TIGGCAATAA AGTITICITAA GATIGAATCC Ø 闰 × O er; \*\*\*\*\*\*\*\*\* щ VAA Noti CCATGGCGGT GTAGAGTACT GGAGCGTTGA Ω O Þ Ü Ø O! Z なり ⋈ E1 U М 2221222 Sphi ပ 4 5 Ø O Ö Þ. ໝ 耳 AGGGGTACGC GAAGCAGATC 4 G D Ħ þ O 771 701 981

111111 Xbai 11111111 Clai

Hindri

gagatgggtt tittatgatta gagtcccgca attatacatt taatacgcga tagaaaacaa aatatagcgc

TACTAGATCG ATAAGCTTCT AGAGCGGCCG GGGACGCGCC ACAACGTCGT CCCTTTCGCC AGCTGGCGTA CCGTCGTTTT AATGGCGAAT CCGCTACACT CGCTCACTGG CGCAGCGTGA CAGCACATCC GCGCAGCCTG TCATCTATGT AATCGCCTTG CCCAACAGTT GTATTACGCG GCGCGCGGTG ATAGTGAGTC GATCGCCCTT TACCCAACTT AATTCGCCCT ATAATTATC ACCCTGGCGT GGCCCGCACC GCAAACTAGG

GGTGGTTACG TTAGTGCTTT TCCTTTCTCG CGGCGGGTGT Tricitacct GGGTTCCGAT GCATTAAGCG CTCCTTTCGC GTGGAGCTCC GACTGGGAAA

GGCCATCGCC GTTCCAAACT GCTCCCTTTA TCACGTAGTG ATAGCGAAGA GGGTCATGGT CTGTAGCGGC CTAGCGCCCG TAAATCGGGG 1331 1401 1471

GTGGACTCTT TTCTTTAATA 1541 1611 1681

ATATTAACGC TTTGCCGATT TTTTAACAAA TATAAGGGAT 1751 1821

CTAAATACAT AGGAAGAGTA TTTTTGCTCA ATATTGAAAA TOCCITCCIO GTTTATITT 1891 1961 2031

GAGTGGGTTA 2101 2171

AGCACTTTTA ATCGTTGGGA CATCGAACTG GCCGCATACA CATGACAGTA ACAACGATCG TCCAATGATG ACTCGCCTTG CAACTCGGTC TTACGGATGG CTTACTTCTG 2241 2311 2381

CTATTCTCAG

TTGCGCAAAC TATTAACTGG CGAACTACTT

GGCAACAACG

CTGTAGCAAT

GGAGCTAACC AATGAAGCCA

GAGGACCGAA ACCEGAGCTG

AGAGAATTAT

GCAGTGCTGC

SUBSTITUTE SHEET (RULE 26)

1191 1261

GGCTGGCTGG GGGCCAGATG GAAATAGACA ATATATACTT AAGGATCTTC AGGGGTGGTT	CAGATACCAA CTACATACCT GTTGGACTCA CCCAGCTTGG TTCCCGAAGG GCTTCCAGGG TTTTTGTGAT TGGCCTTTTG	TACCGCCTTT GAAGCGGAAG ACGACACGCTT GGCACCCCAG CACACAGGAA KPDI TGGGTACCGG	AAAGGTAAGA AAAGGTGGCC GATACGTCAT TCGGTTTTTA ATGCTAATTT TAAAATAGCT AACATTTACA CCGAACCCAAC
CGGCCCTTCC TGCAGCACTG ATGGATGAAC AAGTTTACTC CCTTTTTGAT GAAAGATCA	CAGCAGAGCG GTAGCACCGC GTCTTACCGG GTGCACACAGA AGCGCCACGC GCACGAGGGA TGAGCGTCGA	ATAACCGTAT AGTGAGCGAG TGCAGCTGGC TCACTCATTA ATAACAATTT GAACAAAAGC	AAATAAAAC ATCGGTACTTT TCGGTACTTT TGTATTTGAG AAAAACCCAT ATAATAAGAT TTAGACTCAA CAGCCCAGG
CTTCTGCGCT GCGGTATCAT TCAGGCAACT CTGTCAGACC CTGTCAGACC AGGTCAAAAAAAAACAT	TAACTGGCTT CAAGAACTCT GATAAGTCGT CGGGGGGTTC GCTATGAGA ACAGGAGAGC ACCTCTGACT CGCGGCCTTT	GATTCTGTGG GCAGCGAGTC GATTCATTAA GTGAGTTAGC TTGTGAGCGG TTGTGAAAGG	GGGATAGTCC TAAGTAAAAT GGATGTTTTG AATGCATATC AATATCTTTA AATATCTTTA AATATGAACA AAAGATAAAC CGATCCATAG CTCCACCCC
TGCAGGACCA CGTGGGTCTC CGACGGGGAG GCATTGGTAA AAAAGGGTCT ACTGAGCGTC	TTTCCGAGG GCCACTTT TGCCAGTGGC TCGGGCTGAA TACAGCGTGA CAGGGTCGGA GGGTTTCGCC	GTTATCCCCT ACGACCGAGC CGCGTTGGCC CGCATTAAT TTGTGTGGAA CAATTAACCC	GAAGAGAGTC AAAGGTGGTA TAAAAATTGA GCGATTTGGA ECORI ECORI TGTATAAGA GGCGAATTCC TAGTAAAAATA GGCGAATTCC
CGGATAAAGT AGCCGGTGAG GTTATCTACA CACTGATTAA TTTTTAAATTT GCGTAATCTG	ACCAACTCTT CCGTAGTTAG CAGTGGCTGC GGCGCAGCGG CTGAGATACC CGGTAAGCGG TAGTCCTGTC	TCTTTCCTGC CCGCAGCCGA CCTCTCCCCG GTGAGCGCAA GGCTCGTATG GCCCAAGCGCG	ATATGCTTGA ACATCAGTTA TCTACTATTA AAGTTTATTC ACAGAGGGAT TATATATTTCA GTATTTTTTC AAAGCCCAAA
TGGATGGAGG ATAAATCTGG CCGTATCGTA ATAGGTGCCT TAAAACTTCA TTAAACTTCA TTAACGTGAG	ATCAAGAGCT TCTAGTGTAG ATCCTGTTAC TACCGGATAA CTACACCGAA GACAGGTATC GGTATCTTTA	GCTCACATGT ATACCGCTCG ACGCAAACCG AAAGCGGGCA TTATGCTTCC CCATGATTAC	GAGGTCATTC AAAAGTGAAA TTTACTCTTT ATTGGTAAAT TTTGGAAAAAA GCAGCGATGG CTAAAGTCCT CCACCCCAGT
ATTAATAGAC TTTATTGCTG GTAAGCCCTC GATCGCTGAG TAGATTGATT CCAAAATCCC TTGAGATCCT	TGTTTGCCGG ATACTGTCTA AGACGATAGT AGCGAACGAC GAGAAAGGCG GGAAACGCCT GCTCGTCGG	CTGGCCTITT GCTCACATGT TCTTTCCTGC GTTATCCCCT GATTCTGTGG ATAACCGTAT GAGTGAGCTG ATACCGCTG CCGCAGCCGAGC GCAGCGAGTC AGTGAGCGAG AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGAC GTTCATTAA TGCAGCTGGC TCCCGACTGG AAAGCGGCA GTGAGCGCCA CGCAATTAAT GTGAGTTAGC TCACTCATTA GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA TTGTGAGCGG ATAACAATTT ACAGCTATGA CCATGATTAC GCCAAGCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC XAOI	GCCCCCCTC TTACCTGGTC CAAAGTGAAA TTTTGTATGA AGTTCGTTGC GACATAATTT TGCCCCCGTT AAAACAACCC CCAACCCAAC
2591 2661 2731 2801 2871 3011	3081 3221 3221 3261 3501 3571		44 44 442011 442011 44411 4111 46551

-1G.\_47D

---- MILET IRIII F 261

	TYPTTCTTTT	TCATCGGTAG	TCAGCATTGT GTAGC	TITGAATCCC GAGCITTITT	TCTTCTTTCT TTCTTCTTT TGTGGTAGAA TTTGAATCCC TCAGCATTGT TCATCGGTAG TTTTTCTTTTT CATGATTTGT GACAAATGCA GCCTCGTGC GAGCTTTTTT GTAGCA	TTCTTCTTTT	5181 TCTTCTTTCT TTCTTCTTT TGTGGTAGAA TTTGAATCCC TCAGCS 5251 CATGATTTGT GACAAATGCA GCCTCGTGCG GAGCTTTTTT GTAGC	5181 5251
•	CGGATGTAGA	ATGGGGCTCT	CTCGCGGGGA	CGCCCGGATC	GGCGTGAGTC	9221218299	5111 TCTCGCGGCT GGCGTCTCCG GGCGTGAGTC GGCCCGGATC CTCGCGGGGA ATGGGGCTCT CGGATGTAGA Bglii	5111
H	AGGGGCGGGA Bglii	66766666	CGCCCAGATC	GCGGCTTCGT Bamhi	CTTTGGCCTT GGTAGTTTGG GTGGGCGAGA GCGGCTTCGT CGCCCAGATC GGTGCGCGGG AGGGGCGGGA BANTI BANTI	ggtagtttgg	CTTTGGCCTT	5041
	CICCCCTCC	CTCCTCCCCC	GGACGACGAG CTCCGTTTTT	CTCTTTCTTT	TACCACCACC ACCACCAC COTOCTOCO COTOGOTGOO GGACGACGAG CTCOTOCOCO CTCCCCTOCT GCGGCGCCG GTAACCACCC CGCCCCTCTT CTCCTTT CTCCGTTTTT CTTTTTCGTCT CGGTCTCGAT	GTAACCACCC	GCCGCCGCCG	490T
	GCCCAACCC	ACGAGGCCCG TCCTCCCATC	CCCCCCCTC	TATATACATA	COCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCCAACC	GAAACGCCCC		4831
	ANANCAGCAG	<b>AAGAAAAAGA</b>	AAAGAAAAA	AAAAAAAAG	GCAGCCAAAA	CGCACGTCTC		404

### FIG. 47E

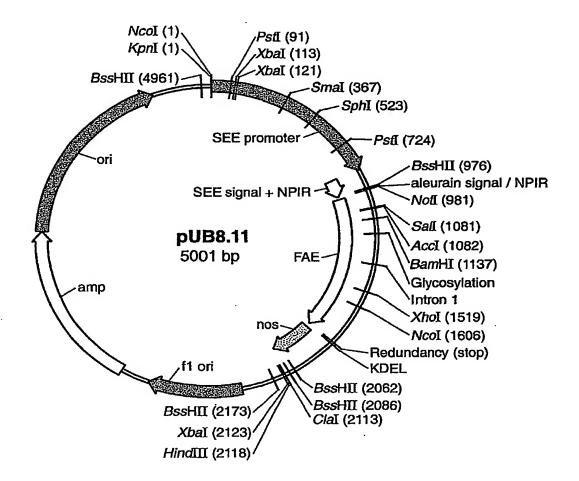


FIG.\_48A

PCTG AGAC

KpbI ~ Camadaaaaa	Kpbl * Candoscon deamanment sections assessant assessant	CHOEKEADOO	4			
GTACCCGGTC	FINCECEGE CATATTAATA CCCTATAGAG TICGTTAATT AGCTITATAG IGGTAACCGA IGTTATA	CCCTATAGAG	TTCGTTTATT	AGCTTTATAG	TGGTAACCGA	ACARTAT TGTTATA
	PstI			XbaI	XbaI	

22222 NCOL

			2		2222222 22222	22222		
71	AGCTCCGAGT	AGCICCGAGT TCTGACTGCA GICTGGATGA CGCGTGTTGT ATCTAGAACT CTAGATAGCA CAGCCACAGG	GTCTGGATGA	CGCGTGTTGT	ATCTAGAACT	CTAGATAGCA	CAGCCACAGC	
	TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAACA TAGATCTTGA GATCTATCGT GTCGGTGTCG	AGACTGACGT	CAGACCTACT	GCGCACAACA	TAGATCTTGA	GATCTATCGT	GTCGGTGTCG	

TCTTTCCTAC CTCCTGACGT GAGGACTGCA AGAAAGGATG CTCTGCCTCG GAGACGGAGC AGTAGTGTTG TCATCACAAC TGTGGACTGT CACGCTGTGA ACACCTGACA GTGCGACACT ACCTACAGGA TGGATGTCCT 141

ATCGTCCCC TAGCAGGGG CCAACAAAT GGTTGTTTA TCACGCGCTC AGTGCGCGAG AGAGTTGGTT TCTCAACCAA ACGGCATCAC TGCCGTAGTG GTCCATTCCA CAGGTAAGGT TGCCGCCGTT ACGGCGGCAA 211

TCCACTGGCC CTTAGAGCGA AGGTGACCGG GAATCTCGCT GTTTTTGTCT CAAAAACAGA CTGTCGCGCC GGAGAGAGAG TACATACATG ATGTATGTAC CCTCTCTCTC ATGTCTTGGC (TACAGAACCG ( 281

### Smal

COTCATGGGA GCAGTACCCT Arcercerce reaccerree TAGCAGCAGC AGTGGGGACC GCTCACTCAT TCAAGATCCC CGAGTGAGTA AGTTCTAGGG GCTCCCGGGA CGAGGGCCCT AATCAGCTCA **LTAGTCGAGT** 351

ATAACCCAAT TATTGGGTTA GTACTGCAAG CATGACGTTC CCCGAACAGA GGGCTTGTCT AGCCTACTCA GTCGGTATAG TCGGATGAGT CAGCCATATC CCTCCGTTGC TGGAAAAGAA ACCTTTTCTT 421

TTTATTTT GAATTAACTG AAATAAAAC CTTAATTGAC TTCATATCGT ACGAAAGCCC AAAACAAACC TTTTGTTTGG AAGTATAGCA TGCTTTCGGG GGTTATCTCT CCAATAGAGA AGTCTAAGGG TCAGATTCCC 491

AATAGGCTGG TTATCCGACC TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AACCTCAACT TACGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCTA TTGGAGTTGA 561

631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT	GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA	
GGGCATTACA	CCCGTAATGT	
ATATCTTTCT	TATAGAAAGA	
CACATCGCAA	GTGTAGCGTT	
TTCTGTAGTG	AAGACATCAC	
TATAGCAAGA	ATATCGTTCT	
CTAATTAATT	GATTAATTAA	
631		

PstI

CGATGAGATG GGTATAAAAC GCTACTCTAC CCATATITTG GARACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG TTCGTTCACC ACTTCGCACC CTTTGTGAGA CGTCTCGGAC TCATCAGCCT AGTAGTCGGA 701

GGACGACCCA TCCCCCTGCC CTCGCCTCGC CCAGTACCAT CTCCCGCCTA GAGGGGGGAT GGGACGCGAG CCCTGCCCTC CCCCGGCACC

771

GCCGCATCCT CTTCTTGGCG CTCGCCGTCT CGGCGTAGGA GAAGAACCGC GAGCGGCAGA CCTGCTGGGT GAGCGGAGCG GTTGCCCACT CGCCGGCGAG ATGGCCCACG CAACGGGTGA GCGCCGCTC TACCGGGTGC GGTCATGGTA CATTTTATGA 990000000 GTAAAATACT 841

11111 Noti Beehil

TEGCCACCEC CECEGTEGEC GCCGCATCNT TEGCGGACTC CAACCCGATC CGGCCGTCA CCGAGCGCGCG ACCGGTGGCG GCGCCACGG CGCCGTAGNA ACCGCCTGAG GTTGGGCTAG GCCGGGCAGT GGCTCGCGCG 911

111111 Noti

GGCCGCCTCC ACGCAGGGCA TCTCCGAAGA CCTCTACAGC CGTTTAGTCG AAATGGCCAC TATCTCCCAA TTTACCGGTG ATAGAGGGTT CCGCCGGAGG TGCGTCCCGT AGAGGCTTCT GGAGATGTCG GCAAATCAGC 981

\*\*\*\*\*\* SalI

AccI

CCGACCTGTG CAACATTCCG TCGACTATTA TCAAGGGAGA GAAAATTTAC AATTCTCAAA GGCTGGACAC GTTGTAAAGGC AGCTGATAAT AGTTCCCTCT CTTTTAAATG TTAAGAGTTT GCTGCCTACG CGACGGATGC 1051

BamHI

HILL	CGGATGGATC CT	CCGCGACG	ACAGCAGCAA	AGAAATAATC	ACCGTCTTCC	1121 CTGACATTAA CGGATGGATC CTCCGCGACG ACAGCAGCAA AGAAATAATC ACCGTCTTCC GTGGCACTGG
AATT	GCCTACCTAG GA	GGCGCTGC	TGTCGTCGTT	TCTTTATTAG	TGGCAGAAGG	CACCGTGACC

ACAATGCAAC TGTTACGTTG ACACCCTACC TGTGGGATGG TGCGGAAAGC ACGCCTTTCG GATGTGGGAG CTACACCCTC TTAGATGTTG AGCTATGATT TCGATACTAA AATCTACAAC ATCACTATGC TAGTGATACG 1191

AGCGAACAGT TCGCTTGTCA GGTTCAGCTC CCAAGTCGAG CCGTCCAGGA GGCAGGTCCT GGATGGGTCT TATAATAA ATATTATAT TACACGGTGG GGTTGTGAAG 1261

CCTACCCAGA ATGTGCCACC CCAACACTIC

CCTCCCTGGC KCCCTCGGCG MGGGAGCCGC GACCGGCCAC CTGGCCGGTG CGCTGACCGT GCGACTGGCA CCGGACTACG GGCCTGATGC TAGCCAGTAT ATCGGTCATA AACAGCAGGT TTGTCGTCCA

GGAGGGACCG Arccecrer Acaccrrces 1331 1401

CGAACCGCGC TAGGCGGACA TGTGGAAGCC ATACGACAAC TATGCTGTTG TGTCTGCGAC ACAGACGCTG CGGCGGGTCG GCCGCCCAGC CCGTGAGTGA GGCACTCACT

ACGACGCAGI GTCGTACATG AACGATGCCT AGCGGCAATC AGGCCTTCGC 1471

TOCTOCOTCA NCOL AACGATGCCT TCCAAGCCTC GAGCCCAGAT TTGCTACGGA AGGTTCGGAG CTCGGGTCTA CAGCATGTAC regeegrad recesaages

ACGCCCATGG TGCGGGTACC CTCGTCCCCA GAGCAGGGGT GCCCCGGTG CGGGGGCCAC TCCCAAACCT TIGCIGCCGT AGGGTTTGGA CACTCATGCC AACGACGGCA GTGAGTACGG ATTTCCGGGT TAAAGGCCCA 1541

GGATGAAGTG CCTACTTCAC TCTGCACTGG AGACGTGACC TTGTGTAAAC AACACATTTG CAGCGCCCAG GICCCGGGIC AACTAGGAAT TTGATCCTTA ATGACCTCGC TACTGGAGCG GCCACATCTC CGGTGTAGAG 1611

ATGACGAGCG TACTGCTCGC AATAAAACCC TTATTTGGG CGCACACGAC CACTTATTAC GCGTGTGCTG GTGAATAATG CGGACAGGGT GCCTGTCCCA AGGCCCAGGG TCCGGGTCCC CAGTGCTGTG GTCACGACAC 1681

TCCTGGAGAG AGGACCTCTC AAGATGGATG TTCTACCTAC GTCATTTCAG CCTCCCGAG TGTACCAGGA CAGTAAAGTC GGAGGGCTC ACATGGTCCT ATGGTGATCA TACCACTAGT GAGCCTGTAC CTCGGACATG 1751

ataaagtitc tatticaaag	GTTAAGCATG CAATTCGTAC	cgcaattata Gcgttaatat	GGTGTCATCT CCACAGTAGA		agtcotatta Tcagcataat	acttaatcgc Tgaattagcg	CCTTCCCAAC GGAAGGGTTG	Gretgerget Cacaccacca	CCCTTCCTTT GGGAAGGAAA
GATCGTTCAA ACATTTGGCA ATAAAGTTTC CTAGCAAGTT TGTAAACCGT TATTTCAAAG	tataattitct gitgaattac atattaaaga caacttaatg	ATTAGAGTCC CGCAATTATA TAATCTCAGG GCGTTAATAT	Bashii TATCGCGCGC ATAGCGCGCG		CTCCAATTCG CCCTATAGTG GAGGTTAAGC GGGATATCAC	CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCTG GCGTTACCCA ACTTAATCGC GACCGGCAGC AAATGTTGC AGCACTGACC CTTTTGGGAC CGCAATGGGT TGAATTAGCG	Caccgatege Gegetageg	CGGCGCATTA AGCGCGGCGG GTGTGGTGGTGGT GCCGCGTAAT TCGCGCCGCC CACACCACCA	GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCCTTCCTTT CACTGGCGAT GTGAACGGTC GCGGGATCGC GGGCGAGGAA AGCGAAAGAA GGGAAGGAAA
		ggttttttagg Ccaaaaatac	TAGGATAAAT ATCCTAITIA			GAAAACCCTG CTTTTGGGAC	AAGAGGCCCG TTCTCCGGGC		CCCGCTCCTT GGGCGAGGAA
AACCACTGAA GGATGAGCTG TAAAGAAGCA TTGGTGACTT CCTACTCGAC ATTTCTTCGT	CGGTCTTGCG ATGATTATCA GCCAGAACGC TACTAATAGT	ATGACGTTAT TTATGAGATG TACTGCAATA AATACTCTAC	BSSHII GCGCGCAAAC CGCGCGTTTG		GCCGGTGGAG	TCGTGACTGG AGCACTGACC	ATCCCCCTTT CGCCAGCTGG CGTAATAGCG TAGGGGGAAA GCGGTCGACC GCATTATCGC	CCTGAATGGC GAATGGGACG CGCCCTGTAG GGACTTACCG CTTACCCTGC GCGGGACATC	CGCCCTAGCG GCGGGATCGC
GGATGAGCTG CCTACTCGAC		ATGACGTTAT TACTGCAATA	acaaaatata Tgittiatat	Xbal	ATCGATAAGC TTCTAGAGCG TAGCTATTCG AAGATCTCGC	ttttacaacg Aaaatgttgc	CGCCAGCTGG GCGGTCGACC	Gaatgggacg Citaccctgc	CACTTGCCAG GTGAACGGTC
AACCACTGAA TTGGTGACTT	ATCCTGTTGC TAGGACAACG	CATGTAATGC GTACATTACG	GCGATAGAAA ACAAAATATA CGCTATCTTT TGTTTTATAT	Clai Bindiii	atcgataagc tagctattcg	CTGGCCGTCG			GTGACCGCTA CACTGGCGAT
GGGGCCGCGT	ttaagaitga Aattctaact	taataattaa Attattaatt	CATTTAATAC GTAAATTATG		atgitactag Tacaatgatc	BSBHII CGCGCGCTCA GCGCGCGAGT	CTTGCAGCAC GAACGTCGTG	agttgcgcag Tcaacgcgtc	TACGCGCAGC ATGCGCGTCG
1821	1891	1961	2031		2101	2171	2241	2311	2381

## -1G.\_48E

2451	CTCGCCACGT		TCGCCGGCTT TCCCCGTCAA AGCGGCCGAA AGGGGCAGTT	GCTCTAAATC CGAGATTTAG	GGGGGCTCCC TTTAGGGTTC CCCCCGAGGG AAATCCCAAG	TTTAGGGTTC AAATCCCAAG	CGATTTAGTG GCTAAATCAC
2521	Ctttacgca Gaaatgccgt	CCTCGACCCC	aaaaacttg Tittitgaac	ATTAGGGTGA TAATCCCACT	TGGTTCACGT ACCAAGTGCA	agtgggccat Tcacccggta	CGCCCTGATA
2591	GACGGTTTT	CGCCCTTTGA GCGGGAAACT	CGTTGGAGTC GCAACCTCAG	CACGTTCTTT GTGCAAGAAA	aatagtggac Ttatcacctg	TCTTGTTCCA AGAACAAGGT	aactggaaca ttgaccttgt
2661	ACACTCAACC TGTGAGTTGG	CTATCTCGGT GATAGAGCCA	Ctattctttt Gataagaaaa	gatittataag Ctaaataitc	ggattttgcc Cctaaaacgg	GATTTCGGCC CTAAAGCCGG	Tattggttaa Ataaccaaft
2731	aaaatgagct Ttttactcga	gattttaacaa Ctaaattgit	aaatttaacg Titaaaitgc		Caaaatatta Gittitataat	acgctttacaa tgcgaatgtt	TTTAGGTGGC AAATCCACCG
2801	actititicgg Tgaaaagccc	GAAATGTGCG CTTTACACGC	CGGAACCCCT	atitgittat Taaacaaata	ttttctaaat Aaaagattta	acaticaaat Tgtaagttta	atgtatccgc Tacataggcg
2871	TCATGAGACA AGTACTCTGT	ataaccctga Tattgggact	taaatgcitc atttacgaag	aataatatig Ttattataac	aaaaaggaag tttttccttc	agtatgagta Tcatactcat	ttcaacattt aagttgtaaa
2941	CCGTGTCGCC	Cttattccct Gaataaggga	TTTTTGCGGC AAAAACGCCG	attttgcctt Taaaacggaa	cctgttttg ggacaaaaac	CTCACCCAGA GAGTGGGTCT	aacgctggtg Ttgcgaccac
3011	aaagtaaaag Tittcatititc	atgctgaaga tacgacttct	TCAGTTGGGT AGTCAACCCA	GCACGAGTGG	GTTACATCGA CAATGTAGCT	actggatete Tgacetagag	aacagcggta Ttgtcgccat
3081	AGATCCTTGA TCTAGGAACT	GAGTTTTCGC CTCAAAAGCG	CCCGAAGAAC GGGCTTCTTG	gtititccaat Caaaaggita	gatgagcact Ctactcgtga	TTTAAAGTTC AAATTTCAAG	TGCTATGTGG ACGATACACC
3151	CGCGGTATTA GCGCCATAAT	TCCCGTATTG AGGGCATAAC	ACGCCGGGCA TGCGGCCCGT		GGTCGCCGCA	TACACTATTC ATGTGATAAG	TCAGAATGAC AGTCTTACTG
3221	TTGGTTGAGT AACCAACTCA	ACTCACCAGT TGAGTGGTCA	CACAGAAAAG CATCTTACGG GTGTCTTTTC GTAGAATGCC		atggcatgac taccgtactg	agtaagagaa Tcattctctt	ttatgcagtg Aatacgtcac

3291	CTGCCATAAC	CATGAGTGAT	AACACTGCGG	CCAACTTACT	TCTGACAACG ATCGGAGGAC	ATCGGAGGAC	CGAAGGAGCT
Ċ	PITHIPPOWS	THOUSE CAROLE		GGTTGAATGA		TAGCCTCCTG	GCTTCCTCGA
7055	TTGGCGAAAA	AACGTGTTGT	ACCCCCTAGE	ACATTGAGCG	CTTGATCGTT GAACTAGCAA	GGGAACCGGA	GCTGAATGAA
3431	GCCATACCAA CGGTATGGTT	acgacgagcg Tgctgctcgc	TGACACCACG ACTGTGGTGC	ATGCCTGTAG TACGGACATC	Caatggcaac Gttaccgttg	aacgitgcgc Tigcaacgcg	aaactattaa Titgataatt
3501	CTGGCGAACT	actitacteta Tgaatgagat	GCTTCCCGGC	aacaattaat Ttgttaatta	agactggatg Tctgacctac	GAGGCGGATA	AAGTTGCAGG TTCAACGTCC
3571	ACCACTTCTG TGGTGAAGAC		CGCTCGGCCC TTCCGGCTGG CTGGTTTATT GCGAGCCGGG AAGGCCGACC GACCAATAA	CTGGTTTATT GACCAAATAA	GCTGATAAAT CGACTATTTA	cregaecege Gaccregece	TCAGCGTGGG
3641	TCTCGCGGTA AGAGCGCCAT	TCATTGCAGC AGTAACGTCG	ACTGGGGCCA TGACCCCGGT	GATGGTAAGC CTACCATTCG	CCTCCCGTAT	CGTAGTTATC	Tacacgacgg Atgtgctgcc
3711	GGAGTCAGGC CCTCAGTCCG	AACTATGGAT TTGATACCTA	gaacgaaata Cttgctttat	GACAGATCGC CTGTCTAGCG	TOAGATAGGT GCCTCACTGA ACTCTATCCA CGGAGTGACT	GCCTCACTGA	ttaagcattg aattcgtaac
3781	GTAACTGTCA CATTGACAGT	GACCAAGTTT CTGGTTCAAA	ACTCATATAT TGAGTATATA	actitegatt Tgaaatceaa	gattttaaac Ctaaattttg	ttcatttta Aagtaaaaat	atttaaaagg taaattttcc
3851	ATCTAGGTGA	agatcctttt Tctaggaaaa	TGATAATCTC ATGACCAAAA ACTATTAGAG TACTGGTTTT	atgaccaaaa tactggtttt	tcccttaacg agggaattgc	TGAGTTTTCG ACTCAAAAGC	ttccactgag Aaggtgactc
3921	CGTCAGACCC GCAGTCTGGG		CGTAGAAAG ATCAAAGGAT CTTCTTGAGA GCATCTTTTC TAGTTTCCTA GAAGAACTCT	CTTCTTGAGA GAAGAACTCT	TCCTTTTTT AGGAAAAAA	CTGCGCGTAA	TCTGCTGCTT
3991	GCAAACAAAA CGTTTGTTTT	AAACCACCGC TTTGGTGGCG	GCAAACAAAA AAACCACCGC TACCAGCGGT GGTTTGTTTG	GGTTTGTTTG	CCGGATCAAG	AGCTACCAAC TCGATGGTTG	TCTTTTCCG AGAAAAAGGC

FIG.\_48G

AAGGTAACTG GCTTCAGCAG AGCGCAGATA CCAAATACTG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTTTATGAC AGGAAGATCA CATCGGCATC AATCCGGTGG

4061

4131	acttcaagaa tgaagttctt	CTCTGTAGCA GAGACATCGT	CCGCCTACAT	ACCTCGCTCT TGGAGCGAGA	GCTAATCCTG CGATTAGGAC	TTACCAGTGG AATGGTCACC	CTGCTGCCAG GACGACGGTC
4201	TGGCGATAAG ACCGCTATTC	TCGTGTCTTA AGCACAGAAT	CCGGGTTGGA GGCCCAACCT	CTCAAGACGA	TAGTTACCGG ATAAGGCGCA ATCAATGGCC TATTCCGCGT		ಕರ್ವಾದಿಕ್ಕರು ರಾವಾಗಿಗಳು
4271	TGAACGGGGG	GTTCGTGCAC CAAGCACGTG	ACAGCCCAGC TGTCGGGTCG	TTGGAGCGAA AACCTCGCTT	CGACCTACAC GCTGGATGTG	CCAACTGAGA GCTTGACTCT	TACCTACAGC ATGGATGTCG
4341	GTGAGCTATG CACTCGATAC	Agaaagcgcc Tctttcgcgg		ACGCTTCCCG AAGGGAGAAA GGCGGACAGG TGCGAAGGGC TTCCCTCTTT CCGCCTGTCC		Tatccggtar Ataggccatt	GCGGCAGGGT CGCCGTCCCA
4411	CGGAACAGGA GCCTTGTCCT	GAGCGCACGA	GGGAGCTTCC CCCTCGAAGG	AGGGGGAAAC TCCCCCTTTG	GCCTGGTATC	titatagecc Aaatatcagg	TGTCGGGTTT ACAGCCCAAA
4481	CGCCACCTCT GCGGTGGAGA	GACTTGAGCG CTGAACTCGC	TCGATTTTTG AGCTAAAAAC	TGATGCTCGT CAGGGGGGG GAGCCTATGG ACTACGAGCA GTCCCCCCGC CTCGGATACC	CAGGGGGGCG	GAGCCTATGG CTCGGATACC	aaaacgcca Ttttgcggt
4551	GCAACGCGGC	CTTTTTACGG GAAAAATGCC		TTCCTGGCCT TTTGCTGGCC AAGGACCGGA AAACGACCGG	TTTTGCTCAC AAAACGAGTG	ATGTTCTTTC TACAAGAAAG	CTGCGTTATC GACGCAATAG
4621	CCCTGATTCT GGGACTAAGA		GTATTACCGC CATAATGGCG	CTTTGAGTGA GAAACTCACT	GCTGATACCG CGACTATGGC	CTCGCCGCAG CCGAACGACC GAGCGGCGTC GGCTTGCTGG	CCGAACGACC
4691	GAGCGCAGCG CTCGCGTCGC	AGTCAGTGAG TCAGTCACTC	CGAGGAAGCG GCTCCTTCGC	GAAGAGCGCC CAATACGCAA CTTCTCGCGG GTTATGCGTT		accecetete Tegeggagag	CCCGCGCGTT
4761	GGCCGATTCA		TGGCACGACA	GGTTTCCCGA	CTGGAAAGCG GACCTTTCGC	GGCAGTGAGC	GCAACGCAAT CGTTGCGTTA
4831	TAATGTGAGT ATTACACTCA	TAATGAGT TAGCTCACTC ATTACACTCA ATCGAGTGAG	ATTAGGCACC TAATCCGTGG	ATTAGGCACC CCAGGCTTTA CACTTTATGC TAATCCGTGG GGTCCGAAAT GTGAAATACG		TTCCGGCTCG	Tatgttgtgt Atacaacaca

# FIG.\_48H

Begill

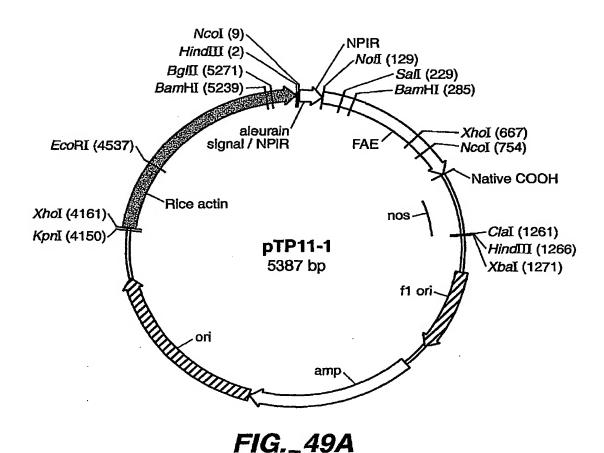
GGAATTGTGA GCGGATAACA ATTTCACACA GGAAACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTTGTCGA TACTGGTACT AATGCGGTTC GCGCGTTAAT

4901

KpnI

ACCCTCACTA AAGGGAACAA AAGCTGGGTA C TGGGAGTGAT TTCCCTTGTT TTCGACCCAT G

4971



SUBSTITUTE SHEET (RULE 26)

NCOL

Hindri

H

AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG >

TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC A Ø Ħ TCGCCTCCTC GCAGGGCATC O 141

Acci

GACCTGTGCA ACATTCCGTC GACTATTATC AAGGGAGAGA AAATTTACAA TTCTCAAACT 0 Ŋ O BendI 211

GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC TCTACAACTC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA Ø V F R G T G E CGTCTTCCGT GGCACTGGTA L G K L L L B B K E I I T CCGCGACGAC AGCAGCAAAAAATAATCAC K H D GCCAGTATCC WILL GATGGATCCT о П O . . 281 351 421 491 561 AFASYMINDAFGABCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA

NGOI

Clai CAATTATACA TTTAATACGC GTGTAGAGTA GTGCTGTGAG GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGA GCCTGTACAT AAGATTGAAT GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGG TGTCATCTAT GTTACTAGAT GGCCGCGTAA ATAATTAACA 闰 ט ರ AAAGTTTCTT GCCCATGGCG CTGGAGAGGG TAAGCATGTA TGCACTGGGG ATGAAGTGCA E E Ħ **☆** ¤ D D IGTAATGCAT GACGTTATTT ATGAGATGGG TTTTTATGAT TAGAGTCCCG TACCAGGAAA GATGGATGTC ATTTGGCAAT GCAGGGGTAC Z TGAATTACGT o Fa O U 7 14 14 14 CCCCGGTGGA CACATTTGTC TCGTTCAAAC TAATTTCTGT **≻ E**4 ρι CTCATGCCAA CGACGGCATC CCAAACCTGC GCGCCCAGAA TCCCCGAGTG AAGAAGCAGA GATTATCATA NN O Z 4 Д Ø GATCCTTACA CATTTCAGCC ATGAGCTGTA GTCTTGCGAT ල ල Ü A A ש Hindiri CTGGAGCGTT GGTGATCAGT CCACTGAAGG CCTGTTGCCG Þ U Ø 0 片 3 701 771 841 981 1051 1121 1191

Clar Xbar

AATGAGCTGA TTTTCGGGGA GGTTCAGTAC CGCGCTCACT TGCAGCACAT TTGCGCAGCC CGCGCAGCGT CGCCACGTTC TTACGGCACC CGGTTTTTCG ACTCAACCCT ATGAGACAAT のゴタゴでほここと AGTAAAAGAT ATCCTTGAGA CGGTATTATC GCCATAACCA TCGTATTACG ATTTAGTGCT TTGGTTAAAA GTATCCGCTC CAGCGGTAAG CTATGTGGCG CTTCCTTTCT CTGGAACAAC TAGGTGGCAC CAACATTTCC AGAATGACTT ATGCAGTGCT TTAATCGCCT TTCCCAACAG GTGGTGGTTA CCCTGATAGA CGCTGGTGAA CTATAGTGAG CACTATTCTC TCTTACGGAT GGCATGACAG TAAGAGAATT GTTACCCAAC TAGGGTTCCG TTGTTCCAAA ATTCAAATAT CCGATCGCCC CGCGGCGGGT GCTTTCTTCC TOGGCCATCG TTTCGGCCTA GCTTACAATT TATGAGTATT CACCCAGAAA TGGATCTCAA TAAAGTTCTG TTCTAAATAC AAAGGAAGAG GGGCTCCCTT TAGTGGACTC AAATATTAAC TGTTTTTGCT CCAATTCGCC AAACCCTGGC GCGCATTAAG CGCTCCTTTC GTTCACGTAG ATTTTGCCGA TACATCGAAC TGAGCACTTT TCGCCGCATA GAGGCCCGCA GTGACTGGGA CCCTGTAGCG TTTATAAGGG CGGTGGAGCT TAATAGCGAA CCCTAGCGCC TCTAAATCGG TAGGGTGATG CGTTCTTTAA AATTTTAACA TTGTTTTT TAATATTGAA THECCEPTCC ACGAGTGGGT TTTCCAATGA AGCAACTCGG CGATAAGCTT CTAGAGCGGC AAAACTTGAT ATTCTTTGA GAACCCCTAT CAGAAAAGCA CCAGCTGGCG ATGGGACGCG CTTGCCAGCG ATTTAACGCG TTTGCGGCAT CGAAGAACGT GCCGGGCAAG TTACAACGTC CCCGTCAAGC TTGGAGTCCA AATGCTTCAA AGTTGGGTGC CCCCTTTCG TGAATGGCGA CCCTTTGACG GACCGCTACA GCCGCCTTTC TCGACCCCAA ATCTCGGTCT AATGTGCGCG TCACCAGTCA GGCCGTCGTT **ITTAACAAAA** AACCCTGATA TATTCCCTTT GCTGAAGATC GTTTTCGCCC CCGTATTGAC 1471 541 1681 1751 891 1401 1611 1821 1961 2031 2101 2171

CCGCTTTTT CATACCAAAC GGCGAACTAC CACTTCTGCG TCGCGGTATC AGTCAGGCAA	CTAGGTGAAG AAACAAAAAA GGTAACTGGC TTCAAGAACT GCGATAAGTC AACGGGGGGT GAACTAGAGAGAGAACAAACAAGAAGAAGAAGAACAAGAAGAAGA	CCACCTCTGA AACGCGGCCT CTGATTCTGT GCGCAGCGAG CCGATTCATT ATGTGAGTTA AATTGTGAGC CCTCACTAAA	TCGGGATAGT TATAAGTAAA GAGGATGTTT GAAATGCATA GAAATATCTT . CCACAATGAA TAAAAGATAA CACGATCCAT
AAGGAGCTAA TGAATGAAGC ACTATTAACT GTTGCAGGAC AGCGTGGGTC CACGACGGGG	TTAAAAGGAT CCACTGACGG TGCTGCTTGC TTTTTCCGAA AGGCCACCAC GCTGCCAGTG GGTCGGGCTG CCTACAGGGTCG	TCGGGTTTCG AAACGCTAGC GCGTTATCCC GAACGCGTTGG ACGCCGTTGG AACGCAATTAAC CGCCAATTAAC	GAGGAGGAG TATAAAAATT TCGCGATTTG ATTTGTATAA ECORI CAGGCGAATT TCTAGTAAAA
CGGAGGACCG GAACCGGAGC CGTTGCGCAA GGCGGATAAA GGAGCCGGTG TAGTTATCTA	CATTITIAAT AGTITITIAAT GCGCGTAATC CTACCAACTC AGCCGTAGTT ACCAGTGGCT AAGGCGCAGC AACTGAGATA TCCGGTAAGC	TATAGETCERG GCCTATGGAA GTTCTTTCCT GGCCGCAGCC CGCCTCTCCC CAGTGAGCGC CAGTGAGCGC CCGGCTCGTAA	TCATATGCTT AAACATCTAT TTAAGTTTAT ATACAGAGGG AATATATTT GGGTATTTTT CTAAAGCCCA
TGACAACGAT TGATCGTTGG ATGGCAACAA ACTGGATGGA TGATAAATCT TCCCGTATCG	TTTAAAACTT CCTTAACGTG CTTTTTTCT GGATCAAGAG CTTCTAGTGT TAATCCTGTT GTTACCGGAT ACCTACACCG	CTGGTATCIT GGGGGGCGA TTGCTCACAT TGATACCGCT ATACGCAAC GGAAAGCGGG CTTTATGCCTT GACCATGATT	TCGAGGTCAT TCGAGGTCAT AATTTACTCT GAATTGGTTT GCTTTTGTAA TTGCAGCGAT CCCTAAAGTC
AACTTACTTC TAACTCGCCT GCCTGTAGCA CAATTAATAG GGTTTATTGC TGGTAAGCCC TGGTAAGCCC	TTTAGATTGA GACCAAAATC TCTTGAGATC TTTGTTTGCC AAATACTGTC CTCGCTCTGC CTCGCTCTGC CTGGAGGATAGG		GGGCCCCCC GATTACCTGG CCCAAAGTGA ATTTTGTAT TAAGTTCGTT TAAGTTCGTT CTTGCCCCCG
CACTGCGGCC GGGGATCATG ACACCACGAT TTCCCGGCAA CCGGCTGGCT TGGGGCCAGA ACGAATAGA	TCATATATAC ATAATCTCAT CAAGGGATCT CCAGCGGTGG CGCAGATACC GCCTACATAC GGGTTGGACT AGCCCAGCTT GCTTCCCGAA	GAGCTTCCAG GATTTTTGTG CCTGGCCTTT ATTACCGCCT AGGAAGCGGA GCACGACAGG TAGGCACCCC TTCACACCCC TTCACACCGC	GCTGGGTACC ACAAAGGTAA TAAAAGGTGG TTGATACGTC AGTCGGTTTT ATATGCTAAT ATAAAATAG
TGAGTGATAA GCACAACATG GACGAGCGTG TTACTCTAGC CTCGGCCCTT ATTGCAGCAC	CCAAGTTTAC ATCCTTTTTG TAGAAAAGAT ACCACGGCTA TTCAGCAGAG CTGTAGCACC GTGTCTTACC TCGTGCACAC	GCGCACGAGG CTTGAGCGTC TTTTACGGTT GGATAACCGT TCAGTGAGCG AATGCAGCTG GCTCACTCAT GGATAACAT	GGGAACAAA CCAAAATAAA ATATCGGTAA TGTCGGTACT TCTGTATTTG TAAAAAACCC CAATAATAAG
2451 2521 2591 2661 2731 2801	22 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	3571 3641 3711 3851 3921 4061	44 44 44 44 44 44 44 44 44 44 44 44 44

FIG. 49D

AGAAAGAAA TACCCCCCC CCGGACGACG TCGTGCGCG GGAGGGGCGG GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG TCGCGAGCAG TTCTCCGTTT GAGCGGCTTC GTCGCCCAGA GITCITCCACCC GAATGGGGCT CICGGATGTA GATCTTCTTT CTTTCTTT TTTGTGGTAG AATTTGAATC CCTCAGCATT GITCATCGGT AGTITITCIT TICATGAITT GIGACAAIG CAGCCICGIG CGGAGCITIT IIGIAGC BanHI ACTATATACA CCCCTCGCTG CTGGCAAATA GCGAGGAGGA AAAAAAAA TCCTCTTTCT TCGCAGCCAA GGCCGGAAAA CCCCATCGCC GTGCAGCCAA CACCTCCTCC CCCCCCCTC AAGAAACGCC CGGTAACCAC ACCCACCCA ACCECACETC GGGTCGTGGG CCACCACCAC TTGGTAGTTT ACCCAACCCA TCCGCTTCCA TGTCCGCACC CCTACCACCA 20220222 AGGTGGGTCC 111111 Bglii GCCCAACCCA CACCGTGAGT CGGCCTCCC TCCCCCCAAC CCCTCCCCCT GARARACAGC TCTCCTCCCA AGCTCCTCCC CCGGCACTAT AAAAGAAAA CGACGAGGCC TTTTTTGT 4831 5181 4761 4901 4971 5041 5251 5321

## FIG.\_ 49E

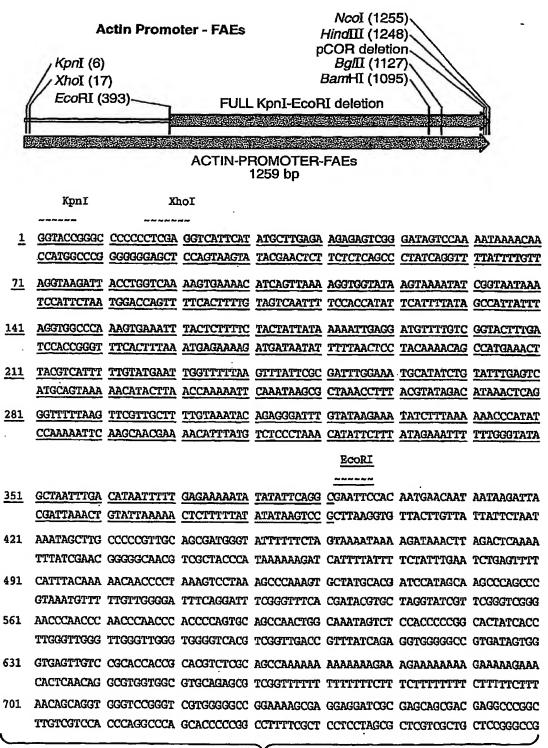


FIG. 50A

SUBSTITUTE SHEET (RULE 26)

771 CCTCCCTCCG CTTCCAAAGA AACGCCCCCC ATCGCCACTA TATACATACC CCCCCTCTC CTCCCATCCC
GGAGGGAGGC GAAGGTTTCT TTGCGGGGGG TAGCGGTGAT ATATGTATGG GGGGGAGAG GAGGGTAGGG

841 CCCAACCCTA CCACCACCAC CACCACCAC TCCTCCCCC TCGCTGCCGG ACGACGAGCT CCTCCCCCCT
GGGTTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCTCCGC CGCCGCCGT AACCACCCCG CCCCTCTCCT CTTTCTTCT CCGTTTTTT TTTCGTCTCG
GGGGAGGCG GCGGCGGCCA TTGGTGGGGC GGGGAGAGGA GAAAGAAAGA GGCAAAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCG CCCAGATCGG TGCGCGGGAG
CAGAGCTAGA AACCGGAACC ATCAAACCCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

### BamHI

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT GGGGCTCTCG CCCGCCCTAG AGCGCCGACC GCAGAGGCCC GCACTCAGCC GGGCCTAGGA GCGCCCCTTA CCCCGAGAGC

BglII

1121 GATGTAGATC TTCTTTCTTT CTTCTTTTG TGGTAGAATT TGAATCCCTC AGCATTGTTC ATCGGTAGIT
CTACATCTAG AAGAAAGAAA GAAGAAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII Ncol

1191 TTTCTTTCA TGATTTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTTTGT AGGTAGAAGC TTACCATGG
AAAGAAAAGT ACTAAACACT GTTTACGTCG GAGCACGCCT CGAAAAAACA TCCATCTTCG AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9, pJQ3.2 and pJO6.3.

FIG.\_50B

### ALEURAIN\_deleted NPIR (Apoplast) Structure and Sequence



ALEURAIN-NPIR-DEL 93 bp

+1 MAHARVLLLALAVLATAAVA Hindlii Nooi

1 AAGCTTACCA TEGCCCACGC COGCGTCCTC CTCCTGGCGC TCGCCGTGCT GGCCACGGCC GCCGTCGCCG
TTCGAATGGT ACCGGGTGCG GGCGCAGGAG GAGGACCGCG AGCGCACGA CCGGTGCCGG CGGCAGCGGC

+1 V A S S R A A

NotI

71 TCGCCTCCTC CCGCGCGGCC GCC AGCGGAGGAG GGCGCGCCGG CGG

FIG.\_51

### SEE1 (Senescence enhanced) PROMOTER sequence

1	CATGGGCCAG	GTATAATTAT	GGGATATCTC	AAGCAAATAA	TCGAAATATC	ACCATTGGCT	ACANTATCTG	
		PstI			XbaI :	XbaI		
		~~~~			~~~~~~	~~~~		
71	AGCTCCGAGT	TCTGACTGCA	GTCTGGATGA	CGCGTGTTGT	ATCTAGAACT	CTAGATAGCA	CAGCCACAGC	
141	ACCTACAGGA	GTGCGACACT	TGTGGACTGT	AGTAGTGTTG	GAGACGGAGC	TCTTTCCTAC		
211	TGCCGCCGTT	GTCCATTCCA	ACGGCATCAC	TCTCAACCAA	TCACGCGCTC	CCAACAAAAT	ATCGTCCCCC	
281	ATGTCTTGGC	GGAGAGAGAG	TACATACATG	CTGTCGCGCC	GTTTTTGTCT	GAATCTCGCT	TCCACTGGCC	
		SmaI						
		~~~~~						
351	AATCAGCTCA	GCTCCCGGGA	GCTCACTCAT	TCAAGATCCC	ATCGTCGTCG	TCACCCCTGG	CGTCATGGGA	
421	TGGAAAAGAA	CCTCCGTTGC	TCGGATGAGT	CAGCCATATC	CCCGAACAGA	GTACTGCAAG	ATAACCCAAT	
			Spl	ıΙ				
			~~~	~~~				
491	TCAGATTCCC	CCAATAGAGA	AAGTATAGCA	TGCTTTCGGG	TTTTGTTTGG	CTTAATTGAC	TTTATTTTTG	
561	TIGGAGTIGA	ATGCTGATTT	GTTGTGTAAA	ATGCCCAACC	ATCTGAATAT	CGAGACGGAT	AATAGGCTGG	
631	CTAATTAATT	TATAGCAAGA	TTCTGTAGTG	CACATCGCAA	ATATCTTTCT	GGGCATTACA	GCTGGAGGCT	
		Ps	tI	•				
All to the first to the first to								
701	TCATCAGCCT	GAAACACTCT	GCAGAGCCTG	AAGCAAGTGG	TGAAGCGTGG	CGATGAGATG	GGTATAAAAC	
771	CCCCGGCACC	GGGACGCGAG	CTCCCGCCTA	CCAGTACCAT	CTCGCCTCGC	TCCCCCTGCC	GGACGACCCA	
841	GTAAAATACT	GTTGCCCACT	CGCCGGCGAG	ATG	•			

### FIG.\_52

### SEE1 (Senescence enhanced) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence

1	CATGGGCCAG	GTATAATTAT PstI	GGGATATCTC	AAGCAAATAA		ACCATTGGCT Chai	ACAATATCTG
71	ልርርጥሮርርልርጥ	ተረጥርኔ (ግርርርኔ	GTCTGGATGA	CCCCTCTTCT	አጥር ጥልር አክርጥ	CTACATACCA	CACCCACACC
141		GTGCGACACT					
211			ACGGCATCAC				
281			TACATACATG			•	
	,	Smal	2300127307110	C10100000	0111110201	Granic 1 cocci	ICCACIGOCC
351	<b>አ</b> ለጥሮክ <i>ር</i> ድርጥሮአ		GCTCACTCAT	መሮአ እርአጥሮርር	N TO COTTO COTTO COTTO	TO A COCO COROCO	COTONICO
421			TCGGATGAGT				
744	HOUNTHON	CCICCGIIGC	TCGGATGAGT Spl		CCCGMACAGM	GTACTGCANG	AIAACCCAAI
			- Di	11			
491	TCAGATTCCC	CCARTAGAGA	AAGTATAGCA	descendante	יוייים איניים	רייויים מיוייויים	Chialalalala) Vilalah
561			GTTGTGTAAA	,			
631			TTCTGTAGTG				
			stI			00001111101	COLGOROGO
			*~~~				
701	TCATCAGCCT	GAAACACTCT	GCAGAGCCTG	AAGCAAGTGG	TGAAGCGTGG	CGATGAGATG	GGTATAAAAC
771			CTCCCGCCTA				
				MAHO	RIL	FLA	LAVL
841	GTAAAATACT	GTTGCCCACT	CGCCGGCGAG	ATGGCCCACG	GCCGCATCCT	CTTCTTGGCG	CTCGCCGTCT
							BssHII
							Noti
	· A T A	A V A	AASI	ADS	NPI	RPV7	ERA
911	TGGCCACCGC NotI	CGCGGTGGCC	GCCGCATCNT	TGGCGGACTC	CAACCCGATC	CGGCCCGTCA	CCGAGCGCGC
	~~~~						
981	· A A GGCCGCC		FIC	G53			

SUBSTITUTE SHEET (RULE 26)

### INTERNATIONAL SEARCH REPORT

Intern Application No PCT/US 01/43588

		J.	101/03 01/43366					
A. CLASS IPC 7	A. CLASSIFICATION OF SUBJECT MATTER IPC 7 C12N15/82							
According to international Patent Classification (IPC) or to both national classification and IPC								
B. FIELDS	SEARCHED							
Minimum de IPC 7	ocumentation searched (classification system followed by classification C12N	on symbols)						
	ion searched other than minimum documentation to the extent that s							
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) WPI Data, EPO—Internal, PAJ, BIOSIS								
C. DOCUM	ENTS CONSIDERED TO BE RELEVANT							
Category °	Cliation of document, with indication, where appropriate, of the rel	evant passages	Relevant to claim No.					
X	US 6 143 543 A (MICHELSEN BIRGIT 7 November 2000 (2000-11-07) see the whole document	1-3, 34-36,74						
Y		4-33, 37-73						
Y	DE VRIES R P ET AL: "The faeA ge Aspergillus niger and Aspergillus tubingensis encode ferulic acid e involved in degradation of comple wall polysaccharides." APPLIED AND ENVIRONMENTAL MICROBI vol. 63, no. 12, December 1997 (1 pages 4638-4644, XP002203731 ISSN: 0099-2240 see the whole document	4-33, 37-73						
		-/						
X Furth	er documents are listed in the continuation of box C.	X Patent family n	nembers are listed in annex.					
Special categories of cited documents:     Thater document published after the international filling date.								
"A" document defining the general state of the art which is not considered to be of particular relevance or priority date and not in conflict with the application but clad to understand the principle or theory underlying the invention of particular relevance; the claimed invention or comment of particular relevance; the claimed invention								
*C* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another catalion or other special reason (as specified)  *C* document referring to an oral disclosure, use, exhibition or cannot be considered novel or cannot be considered novel or cannot be considered to involve an inventive step when the document of particular relevance; the claimed invention cannot be considered novel or cannot be considered novel or cannot be considered to involve an inventive step when the document is combined with one or more other such document.								
other m	eans ni published prior to the international filing date but	thation being obvious to a person skilled of the same patent family						
Date of the a	clual completion of the International search	Date of mailing of t	he international search report					
27	June 2002	12/07/20	·					
Name and m	elling eddress of the ISA  European Patent Ciffce, P.B. 5818 Patentiaan 2  NL - 2280 HV Rijswijk	Authorized officer						
	Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Grosskoj	pf, R					

### INTERNATIONAL SEARCH REPORT

1.3

Inten LApplication No
PCT/US 01/43588

C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT	FC1/03 01,	10000
Category •	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.
А	GARCIA-CONESA MARIA-TERESA ET AL: "A cinnamoyl esterase from Aspergillus niger can break plant cell wall cross-links without release of free diferulic acids." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 266, no. 2, December 1999 (1999-12), pages 644-652, XPO02203732 ISSN: 0014-2956		
		·	
	. 8	- 1	~

### INTERNATIONAL SEARCH REPORT

Interr Application No
PCT/ÜS 01/43588

	agginization on patent taxiny members				PCT/US 01/43588			
Patent document cited in search report		Publication date		Patent family member(s)		Publication date		
US 6143543	А	07-11-2000	GB	230110	3 A ,B	27-11-1996	;	
			*					
-							l	
					•		ĺ	
							ļ	

Form PCT/ISA/210 (patent family ennex) (July 1982)

### This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

### **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

BLACK BORDERS

IMAGE CUT OFF AT TOP, BOTTOM OR SIDES

FADED TEXT OR DRAWING

BLURRED OR ILLEGIBLE TEXT OR DRAWING

SKEWED/SLANTED IMAGES

COLOR OR BLACK AND WHITE PHOTOGRAPHS

GRAY SCALE DOCUMENTS

LINES OR MARKS ON ORIGINAL DOCUMENT

REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY

### IMAGES ARE BEST AVAILABLE COPY.

☐ OTHER:

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.